



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

교육학석사학위논문

중학생의 학업성취도에 대한 사교육의
분위효과 분석과 선택편의 통제

2018년 2월

서울대학교 대학원

교육학과 교육학 전공

김 지 혜

중학생의 학업성취도에 대한 사교육의
분위효과 분석과 선택편의 통제

지도교수 박 현 정

이 논문을 교육학석사 학위논문으로 제출함
2018년 2월

서울대학교 대학원
교육학과 교육학 전공
김 지 혜

김지혜의 석사 학위논문을 인준함
2018년 2월

위 원 장 백 순 근 (인)

부위원장 정 동 욱 (인)

위 원 박 현 정 (인)

국문초록

이 연구의 목적은 분위회귀분석을 이용하여 사교육 효과를 추정할 때 선택편의의 통제에 따라 각 분위 효과의 유의도와 계수가 어떻게 달라지는지 밝히는 데 있다. 이를 위해 서울교육중단연구 6차년도 중학교 3학년 자료를 대상으로, 사교육 참여가 수학과 영어교과의 학업성취도에 미치는 분위별 영향을 추정하였다. 동시에 사교육 참여에 대한 선택편의를 통제한 후, 사교육 참여가 학업성취도에 미치는 영향을 다시 추정함으로써, 선택편의 통제 이전과 이후 사교육 참여의 분위 효과가 어떻게 달라지는지 비교하였다.

이에 따른 연구문제는 아래와 같다.

연구문제 1. 학업성취도에 대한 사교육 참여의 분위효과를 분석할 때, 선택편의를 통제하지 않은 경우 각 분위효과는 어떻게 나타나는가?

연구문제 2. 학업성취도에 대한 사교육 참여의 분위효과를 분석할 때, 선택편의를 통제한 경우 각 분위효과는 어떻게 나타나는가?

이 연구에서는 사교육 참여에 대한 선택편의를 통제하기 위해, 경향점수 역확률 가중치(Inverse Probability Treatment Weight)를 사용하였다. 경향점수 역확률 가중치 방법은 개별 관측치의 배경 변인에 따라 사교육을 받을 확률을 산출하고, 그 확률의 역수를 해당 관측치에 대한 가중치로 사용하는 방법이다. 이 연구에서는 5차년도 자료로부터 얻은 사전 공변인을 사용하여 6차년도 사교육 참여에 대한 경향점수를 구하였으며 한 관측치에 과도하게 가중치가 높게 적용되는 것을 막기 위해 안정화 역확률 가중치를 사용하였다. 다음으로 선택편의 통제에 따른 사교육의 분위 효과 변화를 분석하기 위해, 먼저 조건부 분위회귀분석과 무조건부 분위회귀분석을

각각 실시하고, 각 분포에 대한 분위 효과를 추정하였다. 이어서 안정화 역확률 가중치를 적용한 조건부 분위회귀분석과 무조건부 분위회귀분석을 각각 실시하여 선택편의가 통제된 사교육의 분위 효과를 추정하였다. 이때, 조건부 분위회귀분석에는 학업성취도와 관련이 있는 것으로 알려진 변인 중 6차년도에 수집된 자료를 공변인으로 투입하였다.

연구의 결과와 시사점을 요약하면 다음과 같다.

첫째, 연구의 결과에 따르면 선택편의를 통제하기 전 나타나는 사교육의 분위 효과는 선택편의를 통제한 후 나타나는 분위 효과보다 대체로 더 유의하며 크기 또한 더 높게 추정되었다. 이는 분위회귀분석을 통해 사교육 효과를 추정하는 경우, 선택편의를 통제하지 않은 상태에서는 1종 오류나 과대추정의 가능성이 높아질 수 있다는 점을 시사한다. 즉, 사교육의 효과가 없거나 실제로는 효과의 크기가 낮음에도 불구하고, 사교육 참여에 대한 선택편의를 통제하지 않는 경우 효과가 있는 것으로 드러나거나 효과의 크기가 더 크게 산출될 수 있다는 것을 말한다. 그리고 이는 분위회귀 분석에서도 단순회귀분석과 마찬가지로 처치효과를 산출하기 위해서는 처치집단과 미처치 집단 사이에 무선 할당을 가정하여야 한다는 이론과 상통한다(Abadie, Angrist, & Imbens, 2002; Chernozhukov & Hansen, 2005; Firpo, 2007; Frolich & Melly, 2008).

둘째, 선택편의 통제 전후 사교육 효과의 추정치는 대체로 줄어들거나 유의해지지 않는 방향으로 바뀌었지만 분포의 형태, 분위, 교과에 따라 조금씩 다른 양상을 보였다. 이는 통제 조건 별로, 성적 분위별로, 각 교과별로 선택편의가 다르게 나타난다는 것을 암시한다. 특히, 최상위 분위의 선택편의가 두드러진다는 것을 알 수 있었는데, 이는 대부분의 최상위권 학생들이 참여집단으로 편성됨에 따라 참여집단의 최상위권 점수와 미참여 집단의 최상위권 점수간의 차이가 벌어졌기 때문이다. 또한 수학교과와 경우, 조건부 분포에서는 선택편의 통제가 계수의 유의성에는 영향을 주지 못하였으나, 효과 크기는 분위가 높을수록 크게 감소되었다. 또한 무조건 분포에 대해서는 선택편의를 통제하였을 경우 분위가 높을수록 계수의 크

기가 크게 감소하였고, 최상위권에서 사교육의 효과는 더 이상 유의하지 않은 것으로 나타났다. 한편, 영어교과에서는 수학교과와 다소 다른 양상을 보였는데, 수학 교과의 경우 선택편의의 통제가 주로 상위분위와 하위분위의 효과에 많은 영향을 미치는 것으로 나타난 반면, 영어 교과의 경우 상대적으로 중간에 가까운 분위까지도 유의도 변화에 영향을 미쳤다. 이는 영어 교과의 경우 수학 교과에 비해 더 많은 부분에서 선택편의가 발생하고 있음을 암시한다. 결국, 이러한 결과들로 미루어 볼 때, 사교육 효과 연구에 있어 선택편의는 모든 분포, 분위, 교과에서 동일하게 발생하는 것이 아니라 각각의 세부적인 조건에 따라 다르게 발생한다는 것을 알 수 있다.

주요어 : 사교육 효과, 분위회귀분석, 분위처치효과, 선택편의, 경향점수, 역확률 가중치

학 번 : 2016 - 21503

목 차

I. 서론	1
1. 연구의 필요성 및 목적	1
2. 연구문제	3
II. 이론적 배경	4
1. 사교육 효과 연구 개관	4
2. 사교육의 분위효과 연구	9
3. 분위회귀분석과 선택편의	13
가. 분위회귀분석	13
나. 선택편의	16
다. 분위회귀분석에서의 선택편의	19
III. 연구방법	23
1. 연구대상 및 분석자료	23
2. 연구변인	25
3. 연구모형	30
가. 경향점수 역확률 가중치	30
나. 분위회귀분석모형	31
다. 선택편의가 통제된 분위회귀분석모형	32
IV. 연구결과	33

1. 기술통계	33
2. 경향점수 및 안정화 역확률 가중치 추정	35
3. 분위회귀분석	40
가. 수학교과에 대한 분석결과	40
나. 영어교과에 대한 분석결과	42
4. 선택편의가 통제된 분위회귀분석	44
가. 수학교과에 대한 분석결과	44
나. 영어교과에 대한 분석결과	48
 V. 결론	 53
1. 요약	53
2. 논의	56
 참고문헌	 59
Abstract	64
부 록	68

표 목 차

〈표Ⅲ-1〉 서울교육중단연구 사교육 문항 응답분류	26
〈표Ⅲ-2〉 투입된 변인 설명	28
〈표Ⅳ-1〉 수학교과 주요변인 기술통계	33
〈표Ⅳ-2〉 영어교과 주요변인 기술통계	34
〈표Ⅳ-3〉 수학교과 로지스틱 회귀분석 결과	35
〈표Ⅳ-4〉 영어교과 로지스틱 회귀분석 결과	36
〈표Ⅳ-5〉 역확률 가중치 산출결과	38
〈표Ⅳ-6〉 경향점수 반영 후 수학교과 밸런스 테스트 결과	39
〈표Ⅳ-7〉 경향점수 반영 후 영어교과 밸런스 테스트 결과	39
〈표Ⅳ-8〉 수학교과 조건부 분위회귀분석 결과	40
〈표Ⅳ-9〉 수학교과 무조건부 분위회귀분석 결과	42
〈표Ⅳ-10〉 영어교과 조건부 분위회귀분석 결과	43
〈표Ⅳ-11〉 영어교과 무조건부 분위회귀분석 결과	44
〈표Ⅳ-12〉 수학교과 통제 후 조건부 분위회귀분석 결과	44
〈표Ⅳ-13〉 수학교과 통제 후 무조건부 분위회귀분석 결과	45
〈표Ⅳ-14〉 수학교과 조건부 분위회귀분석 계수 비교	46
〈표Ⅳ-15〉 수학교과 무조건부 분위회귀분석 계수 비교	46
〈표Ⅳ-16〉 영어교과 통제 후 조건부 분위회귀분석 결과	48
〈표Ⅳ-17〉 영어교과 통제 후 무조건부 분위회귀분석 결과	49
〈표Ⅳ-18〉 영어교과 조건부 분위회귀분석 계수 비교	50
〈표Ⅳ-19〉 영어교과 무조건부 분위회귀분석 계수 비교	50

그 립 목 차

[그림IV-1] 수학교과 안정화 역확률 가중치	38
[그림IV-2] 영어교과 안정화 역확률 가중치	38
[그림IV-3] 수학교과 선택편의 통제 전 누적확률분포	47
[그림IV-4] 수학교과 선택편의 통제 후 누적확률분포	48
[그림IV-5] 영어교과 선택편의 통제 전 누적확률분포	51
[그림IV-6] 영어교과 선택편의 통제 후 누적확률분포	51

I. 서론

1. 연구의 필요성 및 목적

사교육은 교육을 비롯한 사회연구에서 가장 빈번하게 등장하는 주제로서 한국 사회 전반에 걸쳐 드러나는 주요 사회적 현안들과 관련되어 있다. 특히 경쟁적 사교육 양상 속에서 파생된 과도한 사교육 참여와 사교육비 지출은 공교육의 내실화를 막고 계층 간 격차를 심화시키는 원인으로 지적되고 있다(남기곤, 2008). 2016년 교육부와 통계청에서 발표한 ‘초·중·고 사교육비조사 보고서’에 의하면 한국의 초·중·고 학생들의 평균 사교육 참여율은 67.8%로 나타났다. 또한 학령인구 감소에도 불구하고 사교육비 총 규모는 18조 1000억 원에 이르러 사교육 참여 학생의 1인당 월평균 사교육비는 37.8만원으로 집계되었다.

정부와 학계는 오래 전부터 이러한 사교육 현상을 한국사회의 커다란 문제로 인식하는 바, 이에 대한 다양한 연구들을 진행해 왔다. 주로 2000년대 이전까지는 사교육 실태나 사교육 참여 요인에 대해 활발한 연구가 진행되었으며 이후 중단연구와 국제학업성취도평가가 도입되면서 사교육의 실질적 효과에 대해 다루는 연구들이 등장하기 시작하였다. 그 중에서 사교육의 효과를 추정하는 연구들은 사교육이 학업성취도 향상에 유의한 영향을 줄 것이라는 일반적인 통념의 사실여부를 확인한다는 점에서 큰 의의를 가진다. 그러나 사교육이 학업성취도에 미치는 영향의 방향과 정도는 일관된 결과를 보이고 있지 않으며, 주로 연구자가 설정한 연구문제나 모형에 따라 상이한 결과가 나타났다.

사실 대부분의 사교육 효과연구들은 사교육의 “평균적 효과(mean effect)”에 관심을 기울여 왔다. 사교육 참여나 사교육비 또는 사교육 시

간과 같은 변수가 한 단위 증가할 때마다 학생들의 평균 학업성취도가 얼마나 증가하는지를 추정하였다. 그러나 사교육의 효과가 꼭 모든 학생들에게 동일하게 나타나는가에 대한 질문이 제기되면서, 일부 연구에서는 사교육의 효과가 학생의 학업성취도 수준에 따라 이질적일 수 있다는 가능성이 드러났다(오영수, 윤정식, 2003; 차동춘, 2015).

학업성취도에 따라 이질적인 사교육의 효과를 분석하기 위해, 학생의 학업성취도를 수준별로 나누어 각 집단에 대한 회귀분석을 실시할 수도 있겠으나, 이는 표본 집단을 임의로 나누는 것이므로 종속변수의 절단(truncation)이 발생할 수 있으며, 회귀식의 추정에 전체 표본을 반영하지 않고 일부 사례만 반영하기 때문에 선택편의의 문제를 가져온다는 점에서 문제가 있다. 따라서 모든 사례수를 고려하지만 가중치를 다르게 주어 분위별 회귀계수를 추정할 수 있는 Koenker와 Bassett(1978)의 분위회귀분석을 이용할 필요가 있다. 실제로 김민선과 백일우(2016), 이광현과 권용재(2011)는 분위회귀분석을 통해 사교육이 학업성취도에 미치는 효과를 분석하였다.

그러나 분위회귀분석으로 추정된 결과를 사교육 처치에 의한 효과로 제시하기 위해서는 단순회귀분석과 마찬가지로 사교육 처치가 무작위로 일어났다는 전제가 있어야 한다(Abadie, Angrist, & Imbens, 2002; Chernozhukov & Hansen, 2005; Firpo, 2007; Frolich & Melly, 2008). 그러므로 만약 처치에 대한 무작위 할당이 일어나지 않았다면 이로부터 얻은 사교육의 효과는 신뢰하기가 어렵다. 이러한 문제의식으로, 이 연구에서는 사교육 참여에 대한 선택편의를 통제하기 전 분위회귀분석을 실시하고, 이후 Firpo(2007)제안한 경향점수 역확률 가중치를 이용해, 선택편의를 통제 한 분위회귀분석의 결과를 실시하여, 선택편의 통제에 따라 사교육의 효과가 어떻게 달라지는지 알아보고자 한다. 또한 선택편의 통제에 따른 사교육 효과의 변화가 분위회귀분석의 선택에 따라, 교과에 따라, 분위에 따라 어떻게 달라지는지 분석함으로써 분위회귀분석의 어떤 조건에서 선택편의의 발생이 두드러지게 나타나는지 알아보고자 한다.

2. 연구 문제

이에 따른 연구의 문제는 아래와 같다.

연구문제 1. 학업성취도에 대한 사교육 참여의 분위효과를 분석할 때, 선택편의를 통제하지 않은 경우 각 분위효과는 어떻게 나타나는가?

연구문제 1-1. 학업성취도에 대한 사교육 참여의 조건부 분위효과를 분석할 때, 선택편의를 통제하지 않은 경우 각 분위효과는 어떻게 나타나는가?

연구문제 1-2. 학업성취도에 대한 사교육 참여의 무조건부 분위효과를 분석할 때, 선택편의를 통제하지 않은 경우 각 분위효과는 어떻게 나타나는가?

연구문제 2. 학업성취도에 대한 사교육 참여의 분위효과를 분석할 때, 선택편의를 통제한 경우 각 분위효과는 어떻게 나타나는가?

연구문제 2-1. 학업성취도에 대한 사교육 참여의 조건부 분위효과를 분석할 때, 선택편의를 통제한 경우 각 분위효과는 어떻게 나타나는가?

연구문제 2-2. 학업성취도에 대한 사교육 참여의 무조건부 분위효과를 분석할 때, 선택편의를 통제한 경우 각 분위효과는 어떻게 나타나는가?

II. 이론적 배경

1. 사교육 효과 연구 개관

우리나라의 사교육 효과에 대한 실증연구는 2000년대 중반 PISA(Program for International Student Assessment)와 TIMSS(Trends in International Mathematics and Science Study)와 같은 국제학업성취도평가와 대규모 종단연구가 도입되면서 본격적으로 이루어지기 시작하였다. 이전에도 사교육의 효과를 밝히기 위한 많은 시도들이 있었지만 표준화된 성취도 데이터 수집에 어려움이 있어, 주로 사교육의 객관적 효과 보다는 학생들이 인지하는 주관적 효과를 분석(윤정일, 1997; 임연기, 1997)하였다. 이어서 2000년대 초반부터 자체 표집을 통해 사교육이 학업성취도에 미치는 객관적 영향을 분석하기 시작하였으며 곧이어 중반부터는 대규모 학업성취도를 활용할 수 있게 되면서 보다 엄밀하고 다각적으로 사교육의 효과를 분석하려는 연구들이 많아졌다. 그런데 이 연구들은 구체적인 연구변인, 분석 방법 등 연구의 설계에 따라 사교육 효과의 유무, 방향, 정도가 상이하다. 따라서 이 장에서는 사교육 효과와 관련된 그동안의 선행연구들을 변인과 분석 방법에 유념하여 정리해 보고자 한다.

먼저 대부분의 선행연구에서는 사교육 효과를 추정하기 위한 종속변수로 학업성취도를 선정하였는데, 이는 일반적인 사교육 참여의 목적이 학생들의 학업성취도 향상에 있다고 보기 때문이다. 이러한 연구들은 대체로 객관적인 학업성취도 점수가 확보되어있는 국내 종단조사나 국제학업성취도 자료에서 제시하고 있는 교과별 수직척도화 점수를 이용하거나, 자체 표집 교내 정기시험 점수나 모의검사 점수를 이용하였다. 이 외에도 수능 점수(김현진, 2007)나 반석차 백분율(박창남, 도종수, 2005)를 학업성취도 변인으로 이용하기도 하였다. 반면 일부 경제학 연구들은 사교육의 효과를

학업성취도가 아닌 대학진학여부나(양정호, 2012) 상위권 대학진학여부(김병철, 2010), 노동시장 진입 후의 월 평균 임금을 통해 분석하기도 하였다(김희삼, 2010).

사교육 효과를 분석하기 위한 독립변수의 선정도 연구에 따라 다소 상이하다. 그러나 대부분의 연구에서는 사교육 참여여부, 시간이나 비용을 채택하고 있다. 초기에는 주로 사교육 참여여부에 따른 학업성취도의 변화만이 주요 관심사였던 반면, 같은 사교육을 받았더라도 사교육의 양이나 질에 따라 그 효과가 달리 나타날 수 있다는 점에서, 점차 시간이나 비용에 따른 학업성취도의 변화를 분석하는 연구들이 등장하였다. 그러나 여전히 연구 방법론이나 문제의 특수성에 따라 참여여부만을 독립변수로 사용하는 연구들이 있으며, 대표적으로는 상경아(2009), 윤유진과 김현철(2016)의 연구가 있다. 한편, 사교육시간에 대한 효과를 분석한 연구로는 김성식과 송혜정(2013), 성기선과 김준엽(2010), 이광현과 권용재(2011)등이 있고, 사교육비용에 대한 효과를 분석한 연구에는 강창희(2012), 박현정(2010), 이수정과 임현정(2009) 등이 있다. 이 외에도 사교육의 지속기간을 주요한 변인으로 판단하여 지속적 과외, 간헐적 과외, 비 과외 여부에 따른 효과를 분석하거나(상경아, 백순근, 2005) 학원수강 여부, 개인과외 여부, 학습지 여부에 따른 효과를 분석한 연구가 있다(임천순, 박소영, 이광호, 2004).

한편 방법론 측면에서 사교육 효과 연구를 살펴보면 다음과 같다. 먼저, 초기에는 주로 단일 시점이나 두 시점에서 사교육 참여집단과 미참여 집단 간의 성취도 평균에 대한 차이 검증을 실시하거나, 중다회귀분석(multiple regression)을 통해 공변량을 통제해주는 방법으로 사교육의 효과를 확인하고자 했다(오영수, 윤정식, 2003; 이수정, 임현정, 2009; 한대동, 성병창, 길임주, 2001).

그러나 위의 연구들이 채택한 방법론은 다음과 같은 한계를 가지는데, 먼저 두 집단 간 평균 차이 검증결과를 과외의 효과로 직결하기가 어렵다는 점이다. 이는 과외 참여 여부 외에도 집단 간 성적 차이에 영향을 미칠 수 있는 배경 변인들이 통제되지 않았기 때문이다. 따라서 만약 두 집단의

차이 검증을 통해 과외 참여의 효과를 밝혀내고 싶다면 두 집단의 기타 변인이 동질한 특성을 가져야 하나, 별도의 통제 없이 구한 두 집단의 점수 차이는 과외 참여 여부에 의한 차이라기보다는 배경 변인의 차이로부터 기인했을 가능성이 있다.

이러한 점에서 다중회귀분석은 사교육 변인 외에도 성취도에 영향을 미칠 수 있는 배경변인을 통제하였다는 점에서 의의가 있으나, 이것 역시 관심변인의 내생성(endogeneity) 문제가 제기될 수 있다. 여기서 내생성이란 다중회귀분석에서 오차항(error term)의 기댓값이, 주어진 모든 독립변수들의 값에서 0이라는 가정이 충족되지 않는 경우를 말한다. 이 경우 관심변인은 ‘내생성(endogeneity)’을 가지며, 반대로 원인변인과 오차항 간에 체계적인 관계가 제거되어 내생성이 없는 경우, 해당 변인은 ‘외생성(exogeneity)’을 갖는다.

이러한 내생성의 원인에 대하여 강창희(2013)는 누락변수(omitted variables), 역의 인과관계(reverse causality), 자기선택(self-selection), 측정 오차(measurement errors)의 네 가지를 제시하였는데 그 중 누락변수와 역의 인과관계의 문제는 사교육 효과분석에서 자주 제기되는 문제이다. 먼저, 누락변수는 결과에 영향을 미칠 수 있는 주요 변인이 모형에 누락된 경우를 말하며 사교육 효과 연구에서는 평균 차이 검정만 실시한다거나 다중 회귀 분석에서 주요한 공변인이 누락된 경우 이 문제가 제기될 수 있다. 또한 원인변인이 결과변인에 영향을 미치는 것이 아니라 반대로 결과변인이 원인변인에 영향을 미치는 경우를 역의 인과관계라고 하는데, 사교육 연구에 있어서 주의해야 할 부분 중 하나이다. 특히, 우리나라의 경우 주로 학업성취도가 높은 학생일수록 사교육에 참여할 확률이 높다고 보고되고 있기 때문에(김성식, 송혜정, 2013; 양정호, 2003) 사교육 효과 연구에서 역의 인과관계의 가능성이 더 높다고 볼 수 있다. 따라서 이에 대한 방법론적 조치가 취해지지 않은 일반적인 중다회귀분석의 결과는 “인과관계” 보다 “상관관계”로 해석하는 것이 적절하다.

이에 따라 최근의 연구에서는 사교육 변인에 대한 내생성을 통제하기

위해 다양한 방법론적 접근을 시도하였는데 대표적으로는 도구변수를 (Instrumental Variable; IV) 이용한 2단계 추정법(Two Stage Least Square; 2SLS)을 활용한 연구(Kang, 2007; 강창희, 2012), 비모수 구간추정법(nonparametric bounds methods)(강창희, 2012), 경향점수매칭(Propensity Score Matching; PSM)(상경아, 2009; 윤유진, 김현철, 2016)이 이용되었다.

먼저 Kang(2007)은 한국교육고용패널(Korea Education Employment Panel; KEEPS)의 일반계 고등학생 자료를 통해 사교육비가 수능 백분위 점수에 미치는 영향을 추정하였다. 그 과정에서 사교육비 지출액에 대한 도구변수로 “첫째아(first-born)” 여부 변수를 사용하였는데 이는 Black et al.(2005)에서 이미 제안된 도구변수로서 개인의 출생순위가 외생적인 변수이며 첫째아가 다른 순위의 아이들보다 더 많은 교육투자를 받는다는 사실에 근거한다. 그 결과 사교육비는 학생들의 학업성취도에 정적으로 유의한 영향을 미치지만 그 크기는 미미한 것으로 나타났다.

강창희(2012)는 중학교 3학년 학생을 대상으로 한 ‘학교교육 수준 및 실태 분석 연구’ 자료를 이용하여 도구변수법과 비모수 구간추정법을 통해 사교육비가 학업성취도에 미치는 효과를 추정하였다. 비모수 구간추정법은 기존 회귀분석에서 가지고 있는 외생성 가정을 완화하는 대신에 점추정치(point estimates) 대신 효과의 하한(lower bound)와 상한(upper bound)을 구하는 방법이다. 그 결과 도구변수법에서 사교육비 지출은 평균적인 학생의 성적을 증가시키는 유의한 효과가 존재하였지만 그 크기는 미미한 것으로 나타났고, 비모수 구간추정법에서 또한 사교육비가 학업성취도에 주요한 영향을 준다는 증거를 발견할 수 없었다.

한편, 상경아(2009)은 한국교육고용패널 자료를 통해 고3 시절 사교육 수강 여부에 따라 수능 수리영역 점수에 차이가 있는지를 경향점수매칭 방법을 활용하여 분석하였다. 먼저 처치변인을 종속변인으로, 공변인들을 독립변인으로 하여 로지스틱 회귀분석을 실시한 후, 산출한 경향점수를 이용해 결합표집을 실시하는데 이 과정에서 학생들이 처치집단에 속할 가능성과 통제 집단에 속할 가능성을 동일하게 만들어 준무선실험 상황을 연

출하였다. 이후 두 집단 간의 평균 수능점수에 대해 t검증을 실시하였는데 그 결과 ‘수리(가)’ 영역에서는 유의미한 차이가 없었지만 ‘수리(나)’ 영역에서 사교육을 받은 집단이 그렇지 않은 집단보다 평균 점수가 높게 나오는 것으로 밝혀졌다.

위의 연구들은 무선헌당이 아닌 상황에서 기존의 방법론이 가지고 있는 처치변인에 대한 내생성 문제를 보완하고자 한 연구들로 다중회귀분석이나 t검증 등의 기존 방법보다 엄격하게 사교육의 효과를 추정한 것이다. 이 연구들은 대체로 사교육의 효과가 유의하지 않거나 유의하더라도 작은 효과 크기를 가지는 것으로 보고하고 있는데 이는 사교육 효과 연구에서 변인에 대한 내생성 통제가 얼마나 중요한지를 암시하는 대목이다.

그러나 위의 방법론들도 다음과 같은 한계를 가지고 있다. 우선, 도구변수 이용법의 경우 정확한 효과 추정을 위해 내생변수와 관련성이 높고 동시에 오차항과는 무관한 도구변수를 선정해야 한다. 그러나 오차항과 무관한 도구변수를 찾기가 쉽지가 않아 Kang(2007)의 연구에서는 ‘처치변인과 오차항의 공분산이 0이상이다’라는 다소 약화된 가정을 채택하여 ‘첫째아’ 변인을 도구변수로 선정하였는데, 근래에는 낮은 출산율로 인해 외동인 경우가 증가하였고, 집안의 재정적 자원이 부족할수록 둘째, 셋째의 출산으로 이어지지 못하는 경우가 많아 첫째 아이에게 사교육 투자를 더 많이 한다는 기존 관계는 위배될 가능성이 있다.

또한 비모수 구간추정법의 경우, 처치변인의 외생성 가정에서 보다 유연할 수 있을지라도 ‘단조적 반응(monotone treatment response; MTR)’과 ‘단조적 선택(monotone treatment selection; MTS)’이라는 강력한 가정이 뒷받침 되어야 한다는 단점이 있다. 여기서 Rosenbaum & Rubin(1983)이 제안한 경향점수를 이용하는 방법이 하나의 대안이 될 수 있는데, 상경아(2009)의 연구처럼 경향점수를 이용하여 매칭을 시행할 수도 있고 역확률가중치 방법(IPTW)을 적용할 수 있다.

2. 사교육의 분위별 효과 연구

사교육 효과를 분석한 기존의 연구들은 주로 사교육의 평균적 효과(mean effect)를 추정한 반면, 사교육의 분포적 효과(distributional effect)를 분석한 연구가 있다(김민선, 백일우, 2016; 이광현, 권용재, 2011). 여기서 평균적 효과라 함은 독립변수가 일정부분 증가할 때 조건부 기댓값(conditional expectation function; CEF)이 증가하는 정도를 말한다. 이를 사교육 연구에서 생각해 보면 다른 공변량을 통제하고 사교육비가 한 단위만큼 증가할 때 기대되는 학업성취도 증가분에 해당한다. 이 값은 일반 선형회귀분석을 통해 구할 수 있으며 반복 기댓값 법칙(law of iterated expectation)이 성립함에 따라 평균정도 학생의 사교육비가 한 단위만큼 증가할 때 기대되는 학업성취도 증가분으로 해석할 수 있다. 평균적 효과는 사교육을 참여함으로써 얻는 효과가 모든 학생들에게 동일하게 발생할 것이라는 ‘동질적 효과의 가정(homogeneous impact assumption)’을 암묵적으로 상정하고 있는데 Heckman, Ichimura와 Todd(1993)는 이 가정의 비현실성을 경험적 분석을 통해 입증한 바 있다(이석원, 2006).

이를 사교육 문제에 비추어 생각해 보면 사교육의 효과가 모든 학생에게 동일하게 나타난다고 가정하는 것인데, 학생의 학업성취도 수준에 따라 사교육 효과가 이질적일 수 있다고 보고한 기존 연구들을(오영수, 윤정식, 2003; 차동춘, 2015) 감안하면 이러한 가정은 무리일 수 있다. 따라서 사교육 효과에 대해 평균적 접근방법만을 취했던 기존과는 다르게, 성취도 수준에 따른 이질적 효과를 분석할 필요가 있다.

또한 이러한 이론적 필요 외에 처방적 관점에서도 성취도 수준에 따른 사교육의 이질적 효과를 분석할 필요가 있다. 어떤 수준의 학생들에게 사교육의 효과가 많이 나타난다는 것은 해당 수준에서 사교육 참여에 따른 교육격차가 크다는 것을 의미하는데, 어떤 수준에서 이런 격차가 많이 벌어지는 지 알 수 있다면, 사교육의 대체제인 EBS나 방과후학교 프로그램

을 개발할 때, 주요 대상을 구체적으로 규정할 수 있어 EBS나 방과후학교의 실효성을 높일 수 있는 배경이 된다.

이러한 관점에서 김민선과 백일우(2016)와 이광현과 권용재(2011)는 Koenker와 Bassett(1978)이 제시한 조건부 분위회귀분석(Conditional Quantile Regression; CQR)을 활용하여 사교육이 학업성취도에 미치는 분위효과를 분석하였다.

조건부 분위회귀분석은 조건부 기댓값대신 조건부 θ 분위(conditional θ -quantile)를 사용하는데 여기서 분위란 0에서 1사이의 값으로 0.25분위의 경우 하위 25%, 0.75분위의 경우 하위 75%, 다시 말해 상위 25%에 해당하는 값을 말한다. 이 분석방법을 이용하면 사교육의 효과를 각 분위별로 추정할 수 있기 때문에 어떤 성적분위의 학생들이 얼마만큼의 사교육 효과를 가지는지 탐색해 볼 수 있다. 물론, 분위회귀분석 대신 학생들을 특정 성적분위로 나누어 각 집단에 대해 일반 회귀분석을 실시할 수도 있다. 그러나 이 경우 학생들의 성적분위를 임의로 선택함에 따라 종속변수의 절단(truncation)이 발생하게 되고 이로부터 구한 추정치는 표본선택편의(sample selection bias)를 갖게 된다(Koenker & Hallock, 2001; Rangvid, 2003) 따라서 모든 사례수를 이용하되 가중치를 다르게 부여하여 계수를 추정하는 분위회귀분석을 실시하여야 한다.

이광현과 권용재(2011)는 사교육 효과 추정에서 가장 먼저 분위회귀분석을 실시한 연구로, 한국교육개발원의 한국교육종단연구(Korea Education Longitudinal Study; KELS) 중학생 1차년도-3차년도 자료를 통해 사교육 효과를 분석하였다. 또한 국어, 영어, 수학교과와 척도 점수를 학업성취도로 두고 사교육 관련 변수로는 사교육시간과 사교육비를 각각 개별 모형으로 투입하였다. 이 연구에서는 성적분위를 0.1, 0.3, 0.5, 0.7, 0.9로 나누어 최하위권, 하위권, 중위권, 상위권, 최상위권에 대한 사교육 시간과 비용의 효과를 분석하였는데 그 결과, 사교육시간은 대부분의 분위에서 유의한 것으로 나타났다. 또한 사교육비는 대부분의 분위에서 정적으로 유의하였으나 0.9 분위에서는 유의하지 않는 것으로 나타났으며 효과 크기 또한 하위

권에서 상위권으로 갈수록 줄어드는 것으로 나타났다. 최상위권에서 사교육 효과가 유의하지 않은 것에 대해서는 천장효과(ceiling effect)가 나타났거나, 최상위권의 사교육 투자의 목적이 현재 성적을 유지하는 데에 있기 때문으로 볼 수 있다고 지적하였다.

한편, 김민선과 백일우(2016)는 사교육이 가장 활발하게 이루어지는 서울시의 데이터를 이용하는 것이 적절하다는 판단 하에 서울교육종단연구(Seoul Education Longitudinal Study of 2010: SELS 2010)의 중학생 데이터를 활용하였다. 이 연구에서는 사교육에 대한 투자가 가장 두드러지는 과목인 영어와 수학과목의 수직척도화 점수를 종속변수로 사용하였으며, 이광현과 권용재(2011)는 이전년도 학업성취도를 고려하지 않았던 데에 반해 바로 직전년도 각 과목 수직척도화 점수를 공변인으로 투입하였다. 또한 사교육비와 사교육시간을 동시에 한 모형으로 투입하여 분석하였다는 점에서 기존 연구와 큰 차이가 있다.

이 연구에서 선정한 성적 분위는 약간의 차이는 있지만 이광현과 권용재(2011)와 마찬가지로 최하위권, 하위권, 중위권, 상위권, 최상위권에 해당하는 0.1, 0.25, 0.5, 0.75, 0.9분위이다. 연구 결과, 영어 사교육시간은 전 분위에서 유의하지 않은데 반해 수학 사교육시간은 0.1분위를 제외한 전 분위에서 유의하였다. 사교육비의 경우 영어와 수학과목 모두 중간 이상의 분위에 해당하는 학생들에게 유의한 것으로 나타났으며 그 값은 분위가 증가할수록 커지는 양상을 보였다. 이는 하위권 일수록 사교육의 효과가 크게 나타났던 이광현과 권용재(2011)의 연구와 다소 상반된 결과인데, 전국데이터가 아닌 서울데이터를 이용하였다는 점, 사교육시간과 비용을 함께 투입한 모형이라는 점이 두 연구 간의 차이를 가져왔을 수 있다.

위의 두 연구 모두 기존의 평균적 효과 대신 사교육의 분포적 효과를 추정하고자 하였다는 점에서 의의가 있지만 다음과 같은 제한점이 있다. 이전의 많은 사교육 효과 연구들이 사교육 변수의 내생성을 통제하기 위해 다양한 방법을 시도하였지만 이 연구들에서는 내생성에 대한 문제가 고려되지 않고 있다는 점이다. 특히 자기선택으로 인한 내생성 문제가 고

려되지 않았기 때문에 이 연구들로부터 추정된 사교육시간과 비용의 효과는 “인과관계” 보다 “상관관계” 일 가능성이 있다. 이 때 발생할 수 있는 편의는 크게 사교육 참여에 대한 선택과 시간과 비용의 정도에 대한 선택으로부터 발생할 수 있는데 사교육을 참여하지 않는 학생들이 전체 빈도 중 약 3분의 1가량을 차지하고 있으며 선행연구들에 의하면 사교육 참여 집단과 미참여 집단은 성취도를 포함한 배경 변수에서 이질적인 성질을 가진다. 따라서 사교육의 분포적 효과를 보다 정확하게 분석하기 위해서는 두 집단의 동질성을 확보할 필요가 있다.

동일한 분야는 아니지만 분위회귀분석 시 발생하는 선택편의를 통제하고자 시도한 연구로는 대표적으로 Firpo(2003), Abadie et al.(1998), Buchinsky(1998, 2001)가 있다. 먼저, Buchinsky(1998, 2001)는 Heckman의 2단계 추정법과 유사한 비모수적 2단계 추정법을 활용하여 선택편의를 통제하여 사업의 분포적 효과를 추정하였다. 또한 Abadie et al.(1998)은 자녀의 성별구성을 도구변수로 활용하여 선택편의를 통제하는 형태의 분위회귀분석을 실시하였다. 마지막으로 Firpo(2007)는 경향점수 가중치를 이용하는 준모수적 추정방법을 제안하였는데 구체적인 과정은 다음과 같다. 먼저, 경향점수추정을 위해 로짓 분석을 하여 처치집단에 참여할 확률을 산출하고 이 값을 가중하여 분위회귀분석을 실시한다. 그의 연구에서는 경향점수의 역수를 가중치로 사용하여 추정한 분위회귀분석이 효율적이라는 것을 이론적으로 검증한 바 있다.

3. 분위회귀분석과 선택편의

가. 분위회귀분석

분위회귀분석이란 Koenker와 Bassett(1978)에 의해 처음 제안된 개념으로, 확률변수의 기댓값이 종속변수인 단순회귀분석과 달리 분위수를 종속변수로 투입하여 회귀식을 산출하는 방법이다.

여기서 분위수(Quantile)란 누적분포함수(Cumulative Distribution Function; CDF)의 값을 τ 이상으로 만드는 확률변수 Y 의 하한(infimum 또는 greatest lower bound)값을 말하며, 정의는 아래와 같다.¹⁾

$$Q_{\tau}(Y) = F^{-1}(\tau) = \inf\{y: P(Y \leq y) \geq \tau\} \quad \langle \text{식 3.1} \rangle$$

이와 비슷한 개념으로 조건부 분위수(Conditional Quantile)도 생각해 볼 수 있다. 조건부 분위수란 일반 분위수와 달리 확률변수 X 의 값이 x 로 주어져 있을 때, 누적분포함수의 값을 τ 이상으로 만드는 확률변수 Y 의 하한 값이며, 이를 식으로 표현하면 아래와 같다.

$$Q_{\tau}(Y|X=x) = F^{-1}(\tau|X=x) \quad \langle \text{식 3.2} \rangle$$

분위수는 사실 기댓값을 분위의 형태로 표현한 것으로, 분위수는 기댓값을 최소화 시키는 문제의 해와 같다(Angrist & Pischke, 2009). 아래는 조건부 분위수에 대한 식을 제시하였다.

$$Q_{\tau}(Y|X) = \operatorname{argmin}_{q(X)} E[\rho_{\tau}(Y - q(X))] \quad \langle \text{식 3.3} \rangle$$

여기서 ρ_{τ} 는 체크함수(check function)를 말하는데 체크함수는 개별 관측치 i 에 대해 Y_i 가 $q(X_i)$ 보다 큰 경우는 양의 가중치를, 반대의 경우 음의

1) 누적분포함수란 X 가 확률변수일 때, $F(x) = P[X \leq x]$ 로 정의되는 함수이다.

가중치를 부여하도록 하는 함수로, 자세한 가중치의 값은 아래에 제시된 체크 함수에 대한 식에 나타난다.

$$\rho_{\tau}(u) = 1(u > 0) \cdot \tau|u| + 1(u \leq 0) \cdot (1 - \tau)|u| \quad \langle \text{식 3.4} \rangle$$

여기서 일반적인 분위회귀분석은 위의 <식 3.5>에서 $q(X)$ 가 선형함수를 따른다는 가정을 하는데 이 가정에 따르면 분위회귀분석의 계수는 아래와 같이 정의된다.

$$\beta_{\tau} = \operatorname{argmin}_b E[\rho_{\tau}(Y - Xb)] \quad \langle \text{식 3.5} \rangle$$

이 때, 만약 관심이 있는 독립변인을 d , 통제를 위한 공변인을 X 라고 하면 위 식은 아래와 같이 다시 쓸 수 있다.

$$(\beta_{\tau}, \delta_{\tau}) = \operatorname{argmin}_{\beta, \delta} E[\rho_{\tau}(Y - \beta X - \delta d)] \quad \langle \text{식 3.6} \rangle$$

한편, $q(X)$ 에 공변인을 투입하지 않고, 오로지 독립변인 d 에 대한 함수를 가정하는 경우에 위 식은 아래와 같이 다시 쓸 수 있다.

$$\delta_{\tau} = \operatorname{argmin}_{\beta, \delta} E[\rho_{\tau}(Y - \delta d)] \quad \langle \text{식 3.7} \rangle$$

여기서 만약 d 가 0 또는 1의 값을 갖는 이분형 변수라면 이는 결국 $q(X)$ 에 어떠한 확률변수도 들어가지 않는 아래와 같은 식이 된다.

$$\delta_{\tau} = \operatorname{argmin}_{\beta, \delta} E[\rho_{\tau}(Y - \delta)] \quad \langle \text{식 3.8} \rangle$$

이 중 $q(X)$ 에 공변인과 관심변인을 함께 투입한 모형을 ‘조건부 분위회귀분석(Conditional Quantile Regression; CQR)’라 하고, 관심변인만을 투입한 모형을 ‘무조건부 분위회귀분석(Unconditional Quantile Regression; UQR)’라 한다.

결국 두 분위회귀분석은 공변인의 투입여부만 다를 뿐, 식으로 표현하는 방법은 동일하다. 사회과학 연구에서는 보편적으로 아래와 같은 방식으로 분위회귀분석의 식을 서술한다.

$$Q_r(Y|X) = \beta_r X$$

〈식 3.9〉

$$Y_i = \beta_r X_i + u_{ri}$$

여기서 오차항인 u_{ri} 는 $q_r(u_i|x_i)=0$ 를 따른다고 가정하지만 단순회귀분석과는 달리 등분산성을 가정하지 않고 있다(Barnes & Hughes, 2002; Rangvid, 2003). 이는 분위회귀분석이 가지는 특징적인 성질이자, 분위회귀분석을 종속변수의 이분산성으로부터 자유롭게 만들어 주는 부분이기도 하다.

한편, 계수를 추정할 때에 단순회귀분석에서는 최소자승법(Ordinary Least Square; OLS)을 이용하여 목적함수의 최솟값을 구하였던 것과 달리, 분위회귀분석에서는 체크함수식에 존재하는 절댓값으로 인해 미분이 불가능한 관계로, 선형 계획법(Linear Programming)이라는 특수한 계산법을 이용한다. 이러한 추정방법의 차이로 인해 분위회귀분석에서는 상대적으로 극단적인 관측치(outlier)의 영향에 덜 민감하다는 장점을 갖고 있다.

한편, 분위회귀분석을 활용할 때는 조건부 분위회귀분석(CQR)을 사용하였는지, 혹은 무조건부 분위회귀분석(UQR)을 실시하였는지에 따라 해석이 달라질 수 있음에 유의하여야 한다. 예를 들어, 무조건부 분포에서 0.9 분위에 해당하는 학생은 전체 표본에서 상위 10%에 드는 학생일지라도, 학생에 대한 배경변인이 통제된 조건부 분포에서는 상위 50%에 해당할 수 있다. 따라서 만약 연구자가 전체의 분포의 상위 성과 집단과 하위 성과 집단에 차별적으로 나타나는 특정 변인의 효과를 보기 위해서는 무조건부 분위회귀모형을 적용하여야 하며 만약 조건부 분위회귀모형을 적용한 경우에는 해당 분위수가 조건부 분포 내에서의 분위수라는 것을 분명히 밝혀야 한다.

나. 선택편의

Rubin(1974)의 인과추론 모형에 따르면, 개인에 대한 처치 효과(treatment effect; TE)를 구하기 위해서는 어떤 개인(i)이 특정 처치(T_i)를 받았을 때 나타날 수 있는 결과(Y_{1i})와 처치에 참여하지 않았을 때 나타나는 결과와의 차이(Y_{0i})를 산출하여야 한다. 이를 처치를 받은 개인에 대해 생각해 보면 i 의 실제 결과와 만약 i 가 처치를 받지 않았다면 얻었을 결과의 차이를 말하며, 반대로 처치를 받지 않은 개인에게 처치 효과란 만약 i 가 처치를 받았다면 얻었을 것으로 예상되는 결과와 실제 결과간의 차이가 된다.

$$\Delta_i = Y_{1i} - Y_{0i} \quad \langle \text{식 3.10} \rangle$$

처치효과 측정 시에 가장 많이 사용되는 통계량(statistics)은 평균 처치효과(Average Treatment Effects, ATE)이다. 여기서 평균 처치효과란 개인별로 이질적인 처치효과의 평균값을 의미한다.²⁾ 이를 식으로 나타내면 아래와 같다.

$$ATE = E[Y_{1i} - Y_{0i}] \quad \langle \text{식 3.11} \rangle$$

여기서 평균 처치효과의 대상이 되는 집단은 전체 집단으로 처치 집단과 미처치 집단의 합집합이다. 만약 처치 집단에 속할 확률을 $P(T_i=1)$ 이라 하고 미처치 집단에 속할 확률을 $P(T_i=0)$ 이라 한다면 평균 처치효과는 조건부 확률의 정의에 의해 <식 3.12>와 같이 나타낼 수 있다.

2) 이 외에도 처치를 받은 사람들의 결과와 이 사람들이 만약 처치를 받지 않았을 시 예상되는 결과와의 차이를 평균한 처치효과(Average Treatment Effects on Treated; ATT)와 처치를 받지 않은 사람들의 결과와 이 사람들이 만약 처치를 받았다면 예상되는 결과와의 차이를 평균한 처치효과(Average Treatment Effects on Untreated; ATU)도 사용되기도 한다.

$$ATE = P(T_i = 1)E[Y_{1i} - Y_{0i} | T_i = 1] + P(T_i = 0)E[Y_{1i} - Y_{0i} | T_i = 0]$$

<식 3.12>

그러나 우리가 현실에서 관찰할 수 있는 결과는 처치를 받은 개인과 처치를 받지 않은 개인의 실현된 결과(realized outcome)(Y_i)일 뿐이다. 예를 들어 한 개인이 이미 처치를 받았다면 처치를 받은 경우의 결과와 처치를 받지 않은 경우의 결과는 동시에 존재할 수 없다. 오직 처치를 받은 개인이 처치를 받아 나타낸 결과만 존재한다. 이는 처치를 받지 않은 개인도 마찬가지이다. 처치를 받지 않은 개인은 이미 처치를 받지 않은 상태에서 처치를 받지 않은 결과만 알 수 있다. 결국 우리가 처치효과 추정을 위해 알고자 했던 Y_{1i} 와 Y_{0i} 는 실제로는 관찰될 수 없는 값이다. 따라서 Y_{1i} 와 Y_{0i} 를 잠재적 결과(potential outcome) 또는 실현된 결과에 대한 각각의 가상적 대응치(counterfactual)라고 지칭한다.

$$Y_i = \begin{cases} Y_{1i} & \text{if } T_i = 1 \\ Y_{0i} & \text{if } T_i = 0 \end{cases} \quad \text{<식 3.13>}$$

$$= Y_{0i} T_i + Y_{1i} (1 - T_i)$$

일반적인 상황에서 잠재적 결과는 직접 산출할 수 없지만 만약 처치가 모든 사람들을 대상으로 무작위로 일어났다는 가정이 충족된다면 평균 처치효과는 <식 3.14>와 같은 방법으로 구할 수 있다.

$$\begin{aligned} & E[Y_{1i} | T_i = 1] - E[Y_{0i} | T_i = 0] \quad \text{<식 3.14>} \\ &= E[Y_{1i} | T_i = 1] - E[Y_{0i} | T_i = 1] \\ &= E[Y_{1i} - Y_{0i} | T_i = 1] \\ &= E[Y_{1i} - Y_{0i}] \end{aligned}$$

Rosenbaum과 Rubin(1983)은 이를 ‘강한 무관성 가정(Strong Ignorability)’로 정의하였는데 이 가정은 아래 두 조건을 만족해야 성립할 수 있다. 첫 번째는 처치에 대한 조건부 독립성 가정(Unconfoundedness

Assumption; Conditional Independence Assumption)으로 공변인이 주어졌을 때 처치는 결과변인과 서로 독립이어야 한다는 가정이다. 두 번째는 공통 영역의 가정(Common Support Assumption)으로 주어진 공변인에서 처치집단과 미처치집단의 처치확률은 어느 정도 중복되는 분포를 가지고 있어야 한다는 가정이다. 이를 식으로 표현하면 아래와 같다.

$$(1) (Y_{0i}, Y_{1i}) \perp\!\!\!\perp T_i | X \quad (\text{조건부 독립성 가정}) \quad \langle \text{식 3.15} \rangle$$

$$(2) 0 < P(T=1|X) < 1 \quad (\text{공통 영역의 가정}) \quad \langle \text{식 3.16} \rangle$$

그러나 만약 처치가 무작위로 일어나지 않아 강한 무관성의 가정을 충족하지 않는다면 <식 3.17>의 두 번째 등호는 더 이상 성립되지 않으며 이를 새로운 식으로 정리하면 다음과 같다.

$$\begin{aligned} & E[Y_i | T_i = 1] - E[Y_i | T_i = 0] \quad \langle \text{식 3.17} \rangle \\ &= E[Y_{1i} | T_i = 1] - E[Y_{0i} | T_i = 1] \\ &+ E[Y_{0i} | T_i = 1] - E[Y_{0i} | T_i = 0] \end{aligned}$$

즉, 처치가 무작위로 일어나지 않은 경우 평균 처치효과를 구하기 위해 처치집단과 미처치 집단의 평균값 차를 구하게 되면 이 값은 평균 처치효과가 아닌 ATT 와 $E[Y_{0i} | T_i = 1] - E[Y_{0i} | T_i = 0]$ 를 합한 결과가 된다. 여기서 $E[Y_{0i} | T_i = 1] - E[Y_{0i} | T_i = 0]$ 를 처치가 무작위성을 띄지 않음으로써 발생하는 선택편의(selection bias)라 일컫는다(Angrist & Pischke, 2009).

선택편의가 발생하는 것을 막을 수 있는 가장 좋은 방법은 무작위 할당 실험을 통해 처치와 결과변인간의 독립성을 확보하는 것이다. 그러나 사람을 연구대상으로 하는 대부분의 사회과학연구에서 실험을 기획하여 실시하기까지는 많은 제재가 뒤따른다. 이에 대한 대안으로 여러 가지 준실험적 방법이 제안되었는데 대표적으로 경향점수를 이용한 방법이 있으며 이 방법은 많은 공변인을 고려할 수 있으면서 동시에 우수한 간명성(parsimony)을 유지할 수 있다(Rajeev & Sadek, 2002). 경향점수를 구하는 방법에 대해서는 연구방법 및 모형에서 자세하게 서술하였다.

다. 분위회귀분석에서의 선택편의

평균처치효과의 정의가 개인에게 나타나는 이질적인 처치효과의 평균임과 동시에 처치여부에 따른 각 집단의 평균 차이임을 생각해 볼 때, 평균 처치효과가 나타내는 정보는 다소 제한적일 수 있다. 예를 들어, 어떤 처치를 평가할 때, 결과변인의 상대적 위치에 따른 처치효과를 각각 구할 수 있다면 이는 평균 처치효과보다 훨씬 풍부한 정보가 될 수 있다. 이를 사교육 문제에 대입해보면 사교육의 평균적 효과만 제시하기보다 학업성취도의 상대적 위치에 따라 이질적인 효과에 대한 정보를 제공하는 것이 사교육이 학생들의 학업성취도에 미치는 영향에 대한 훨씬 다각적인 정보를 주는 것과 같다.

이를 위해 가장 먼저 생각해 낼 수 있는 방법은 학생들을 성적에 따라 최상위권, 상위권, 하위권, 최하위권 등으로 분류한 후, 분류된 각 집단에 대하여 t검증 또는 회귀분석을 실시하는 것인데, 이 경우 표본이 종속변인의 값에 따라 분리되기 때문에 표본선택(sample selection)의 문제가 발생할 수 있다(강창희 외, 2013; 길혜지, 백순근 2016). 따라서 모든 표본을 분석의 대상으로 포함하되 가중치를 다르게 주는 Koenker와 Bassett(1978)의 분위회귀분석이 적절한 대안이 될 수 있다. 그러나 분위회귀분석 또한 마찬가지로 처치가 무작위로 할당되었다는 가정이 충족되어야 여기서 나온 결과를 처치효과로 받아들일 수 있다.

이런 맥락에서 나온 개념이 Doksum(1974)과 Lehman(1974)에 의해 제안된 분위처치효과 (Quantile Treatment Effect; QTE)인데, 이 때 분위처치효과란 고정된 퍼센타일(percentile)에 대한 두 개의 누적 분포함수(Cumulative Distribution Function; CDF) 사이의 거리를 말한다(Firpo, 2007).³⁾ 여기서 비교대상이 되는 누적 분포 함수가 두 개인 이유는 전체의

3) 통계량은 평균처치효과와 마찬가지로 전체에 대한 분위처치효과(Quantile Treatment Effect; QTE)와 처치 집단에 대한 분위처치효과(Quantile Treatment Effect on the Treated; QTT)로 구분할 수 있다.

집단이 처치 또는 미처치로 나뉜다고 보았기 때문이다.

분위처치효과(Δ_τ)를 수식으로 정의하면 다음과 같다. 만일 어떤 개인(i)이 무작위로 특정 처치(T_i)를 받았을 때 나타날 수 있는 결과를 Y_{1i} , 처치를 받지 않았을 때 나타날 수 있는 결과를 Y_{0i} 라 할 때 아래와 같이 표현할 수 있다.

$$\Delta_\tau = Q_{1,\tau} - Q_{0,\tau} \quad \langle \text{식 3.18} \rangle$$

그런데 평균처치효과에서 사용하였던 기댓값과 달리 분위연산자(Quantile operator)는 선형성을 가지고 있지 않다. 따라서 평균처치효과에서는 개인의 이질적인 효과들의 기댓값이 처치집단 결과변수의 기댓값을 미처치집단 결과변수 기댓값과 뺀 값과 같았으나⁴⁾, 분위 연산자에서는 비슷한 연산이 항상 성립되지 않는다.

$$\begin{aligned} Q_\tau(\hat{\beta}) &= Q_\tau(Y_{1i}) - q_\tau(Y_{0i}) \\ &\neq Q_\tau(Y_{1i} - Y_{0i}) \end{aligned} \quad \langle \text{식 3.19} \rangle$$

해당 연산이 성립하기 위해서는 순위 보존성 가정(rank preservation assumption)이 추가로 필요한데, 여기서 순위 보존성 가정이란 처치 여부와 상관없이 집단 내에 있는 개인의 순위는 같다는 것을 말한다. 이 가정이 성립하기 어려운 경우, 분위회귀로부터 얻은 추정치를 동일한 개인에 대한 처치효과로 해석하기 보다는, 각 집단의 분포에 대한 처치효과로 해석하는 것이 적절하다. 즉, τ 분위에 있는 어떤 사람이 처치를 받았을 때와 받지 않았을 때 나타나는 성과의 차이로 해석하는 것이 아니라(강창희, 2013) 처치 집단에 있는 τ 분위의 성과 변수와 미처치 집단에 있는 τ 분위의 성과 변수 값의 차이로 해석하여야 한다고 할 수 있다.

한편, 분위처치효과에서도 평균처치효과와 마찬가지로 처치와 종속변인의 결과를 인과관계로 해석하기 위해 강한 무관성 가정(Rosenbaum and

4) 기댓값은 선형성을 가지는 연산자로, $E[Y_{1i} - Y_{0i}] = E[Y_{1i}] - E[Y_{0i}]$ 가 성립한다.

Rubin, 1983)이 필요하다. 때문에 처치효과 추정치로써 분위회귀분석의 결과를 제시하기 위해서는 처치여부가 무작위인 실험 데이터를 이용하여야 하며 관찰 데이터를 이용할 경우 강한 무관성 가정을 충족시켜 줄 통계적 조치가 필요하다.

이러한 맥락에서 분위회귀분석에서 선택편의를 통제하고자 시도한 연구들이 있다(Abadie et al., 2002; Buchinsky, 1998, 2001; Chernozhukov & Hansen, 2005; Firpo, 2007; Frolich & Melly, 2008). 대표적으로 Abadie et al.,(2002), Chernozhukov와 Hansen(2005)은 이분형 도구변수를 이용하여 조건부 분포에 대한 분위처치효과를 구하는 방법을 제안하였고, Frolich와 Melly(2008)는 같은 방법을 이용하되, 무조건부 분포에 대한 분위처치효과를 구하는 방법을 제안하였다. 또한 Firpo(2007)는 경향점수를 이용한 2단계 접근으로 분위처치효과를 구하는 방법을 제안하였는데 이 방법은 먼저, 처치 변인을 종속변수로 하는 로지스틱 회귀분석을 실시하여 처치에 대한 경향점수를 산출한 후, 산출된 경향점수의 역수를 개별 관측치에 대한 가중치로 사용하여 분위회귀분석을 실시하는 방법이다(Liao & Zhao, 2017).

Firpo(2007)에 따르면 강한 무관성 가정이 성립할 때, 아래와 같은 정리가 성립한다.

$$\begin{aligned}
 \tau &= P[Y_1 \leq Q_{1,\tau}] \\
 &= E[P[Y_1 \leq Q_{1,\tau}|X]] \\
 &= E[P[Y_1 \leq Q_{1,\tau}|X, T=1]] \\
 &= E[P[Y \leq Q_{1,\tau}|X, T=1]] \\
 &= E[E[T1\{Y \leq Q_{1,\tau}\}|X, T=1]] \\
 &= E\left[\frac{1}{p(X)}E[T1\{Y \leq Q_{1,\tau}\}|X]\right]
 \end{aligned}$$

$$= E\left[\frac{T}{p(X)} 1\{Y \leq Q_{1,\tau}\}\right] \quad \langle \text{식 3.20} \rangle$$

첫 번째 등식은 분위수의 정의에 의해 성립하며, 두 번째 등식은 반복 기댓값 법칙(Law of Iterated Expectation)에 의해 성립한다. 또한 세 번째 등식은 강한 무관성 정리에 의해, 네 번째 등식은 Y 의 정의⁵⁾에 의해 성립한다. 다음으로 다섯 번째 등식은 $E[1\{A\}] = P(A)$ 라는 사실로부터 성립되고, 여섯 번째 등식은 Z 가 확률변수 일 때, $\langle \text{식 3.21} \rangle$ 에 의해 성립된다. 마지막으로 일곱 번째 등식은 반복 기댓값 법칙에 의해 성립하게 된다.

$$E[Z|X] = P(X)E[Z|X, T=1] + (1-P(X))E[Z|X, T=0] \quad \langle \text{식 3.21} \rangle$$

같은 방법으로 $Q_{0,\tau}$ 를 전개하면 $Q_{0,\tau} = \tau = E\left[\frac{1-T}{1-p(X)} 1\{Y \leq Q_{0,\tau}\}\right]$ 의 음함수가 된다. 이때의 분위처리효과는 $\hat{\Delta}_\tau = \hat{Q}_{1,\tau} - \hat{Q}_{0,\tau}$ 이며, 가중치 $\hat{w}_{0,i}$ 와 $\hat{w}_{1,i}$ 는 $\langle \text{식 3.24} \rangle$ 와 같다.

$$\hat{Q}_{j,\tau} \equiv \underset{\Delta}{\operatorname{argmin}} \sum_{i=1}^N \hat{w}_{j,i} \rho_\tau(Y_i - \Delta) \quad \langle \text{식 3.23} \rangle$$

$$\hat{w}_{1,\tau} = \frac{T_i}{N \hat{p}(X_i)}, \quad \hat{w}_{0,\tau} = \frac{1-T_i}{N(1-\hat{p}(X_i))} \quad \langle \text{식 3.24} \rangle$$

또한 기존의 무조건부 분위회귀분석에 대해서만 처리변인의 효과를 추정하였던 Firpo(2007)의 방법을 Liao와 Zhao(2017)는 조건부 분위회귀분석에 대해서도 적용하여 처리변인의 효과를 추정하였다.

5) $Y = TY_1 + (1-T)Y_0$

Ⅲ. 연구방법

1. 연구대상 및 분석자료

이 연구에서 이용하는 분석 자료는 서울특별시 교육연구정보원에서 수집한 서울교육종단연구(Seoul Education Longitudinal Study of 2010; SELS 2010) 6차년도 중학생 데이터이다. SELS 2010은 2010년 당시 초등학교 4학년, 중학교 1학년, 일반계고 1학년 및 전문계고 1학년을 대상으로 표본을 선정한 후 추적 조사한 패널자료로, 현재 6차년도 자료까지 공개하고 있다(박현정, 하여진, 박민호, 오현철, 이준호, 이진실, 2011). 이 표본은 층화2단계 집락추출을 통해 수집된 것으로 중학교의 경우, 먼저 서울특별시 내 지역교육 지원청을 층으로 선정하고, 주어진 층 내에서 학교를 무선추출한 후 표집된 학교 내에서 다시 2개 학급을 무선 추출하는 방식으로 구성되었다. 연구의 대상이 되는 학생들은 1차년도 당시 초등학교 4학년 패널로 표집된 학생들로서 6차년도를 기준으로 중학교 3학년에 해당하는 319개 학교 3,673명이다. 이 중 수학과 영어교과를 기준으로 종속변수와 처치변수 중 하나라도 결측이 있는 경우는 사례 제거(listwise)하였다. 이 때, 결측은 교과별로 다르게 나타나기 때문에 이 연구에서 최종적으로 분석 대상이 되는 학생의 수는 수학교과와 영어교과와 각각 3,289명, 3,251명이다.

이 연구에서 해당 자료를 분석 대상으로 선정한 이유는 다음과 같다. 우선, 전국단위의 데이터의 경우 읍, 면지역에서는 학원이나 과외에 참여하기 위한 충분한 여건 또는 의지가 있음에도 불구하고 공급이 마땅치 않아 사교육에 참여하지 않을 가능성이 있다. 또한 참여한다 하더라도 원격으로 수강할 수 있는 인터넷 강의 등에 참여함으로써 사교육을 생각할 때

일반적으로 다루는 학원 및 과외와 성격이 다를 수 있다. 그러나 이 연구에서는 사교육의 유형을 구분하지 않고 사교육의 참여, 미참여 여부만 처치변수로 투입하고 있으므로 상대적으로 비슷한 환경에 놓여있고 사교육 참여집단의 사교육 유형이 상대적으로 동질한 서울지역을 대상으로 분석하는 것이 적절하다.

또한 초등학교의 경우 2016년 기준, 가장 높은 사교육 참여율을 보이고 있으나, 전체 사교육 참여비율 대비 일반교과 참여비율의 비는 초, 중, 고에서 가장 낮기 때문에, 사교육에 참여한다고 하더라도 사교육 참여의 주된 목적이 일반 교과의 학업성취도 향상에 있다고 보기 어렵다.⁶⁾ 또한 고등학교의 경우 상급학년으로 진학할수록 사교육 참여율이 현저히 낮아지고, 일부 특수목적 고등학교(이하 특목고)나 자립형사립학교의 경우 사교육에 참여하지 않고 학교에서 제공하는 프로그램에 참여하는 경우들이 많아 분석 대상으로 선정하지 않았다. 반면에 중학교의 경우, 일반 교과에 대한 사교육 참여율이 가장 높고, 1인당 월평균 사교육비도 가장 높아 사교육 참여가 가장 과열되게 나타나는 시기라고 보아 분석의 대상으로 선정하였다.

한편, 중학교 시기의 사교육 효과를 추정하기 위해 이 연구에서는 분석대상을 중학교 3학년으로 선정하였는데 이는 조사시점 당시 자유학기제가 부분적으로 시행되고 있었기 때문이다. 서울교육종단연구는 매년 7월에 조사를 실시하고, 사교육 참여 문항의 경우 조사시점으로부터 직전 1년간 사교육 참여 여부에 대해 묻고 있다. 즉, 중학교 2학년에게 사교육 참여 여부를 묻는 경우, 응답자는 조사시점보다 1년 이전인 중학교 1학년 여름방학부터 중학교 2학년 여름까지 해당사항에 대해 응답하여야 한다. 그러나 당시 자유학기제가 학교에 따라 선택적으로 시행됨에 따라 중학교 1학년 시기의 자유학기제 선택여부가 오염변인으로 작용할 소지가 있어, 자유학

6) 2016년 기준, 초, 중, 고등학교의 사교육 참여율은 각각 80.0%, 63.8%, 52.4%로 나타났으며, 일반교과 사교육 참여율은 각각 53.5%, 55.8%, 43.2%로 나타났다(통계청, 2017). 이에 따라 산출한 전체 사교육 참여율 대비 일반교과 사교육 참여율의 비율은 초, 중, 고등학교가 각각 66.9%, 87.5%, 82.4%이다.

기제 시행 시점으로부터 비교적 시간이 많이 지난 중학교 3학년을 대상 학년으로 선정하여 연구를 진행하였다.

2. 연구변인

이 연구에서 이용하는 종속변인은 수학, 영어교과의 6차년도 수직척도 점수이다. 여러 교과들 중 수학과 영어를 관심교과로 선택한 이유는 2016년을 기준으로 수학과 영어 교과에서 가장 높은 사교육 참여율과 사교육비 지출 양상을 보이고 있기 때문이다⁷⁾(교육부, 2017). 이어서 처치변인은 조사 시점을 기준으로 직전 1년 동안 교과별 사교육 참여 여부로 학부모 조사지에 대한 반응을 이용하였다. 이 조사지의 문항은 먼저 모든 교과에 대한 사교육 참여 여부를 묻고 이후 세부 문항을 통해 각 교과에 대한 사교육 참여 여부를 묻는 순서로 설계되었으나 구조적 결측 반응이 따로 코딩되어 있지 않은 관계로 <표 III-1>과 같이 참여 집단과 미참여 집단을 새롭게 분류하였다. 즉, 교과별 응답을 우선순위로 하되 이전 문항에서 ‘미참여’라고 응답하더라도 교과별 응답에 ‘참여’라 응답하면 ‘참여’ 집단으로 분류하였다.

7) 학교급을 나누지 않고 산출한 과목별 사교육 참여율은 수학과 영어교과가 각각 42.3%와 39.1%, 중학생의 교과별 1인당 월평균 사교육비는 수학이 10.8만원, 영어가 10.1만원으로 집계되었다.

<표 III-1> 서울교육중단연구 사교육 문항 응답분류

전체 응답	교과별 응답	수학(N)	영어(N)	분류
참여	참여	2,474	2,323	참여
	미참여	131	245	미참여
	무응답	71	108	무응답
미참여	참여	28	28	참여
	미참여	79	80	미참여
	무응답	660	659	미참여
무응답(-9)	참여	107	103	참여
	미참여	44	47	미참여
	무응답	50	51	무응답

한편, 조건부 분위회귀분석에서 사용하는 8개의 공변인들은 6차년도 자료에서 수집된 자료이며 이론적 배경에서 다룬 선행연구들을 토대로 선정된 변인들이다. 여기서 수업태도, 교사에 대한 평가, 학습 방법 변인은 리커트 척도 문항으로써 학생들은 각 문항에 대해 ‘전혀 그렇지 않다(1)’, ‘그렇지 않다(2)’, ‘보통이다(3)’, ‘그렇다(4)’, ‘매우 그렇다(5)’로 응답할 수 있다. 이 변인들은 같은 구인을 측정하는 복수의 세부 문항으로 구성되어 있으며 이들의 평균값을 산출해 해당 변인을 구성하였다. 또한 특목고 진학희망 변인은 본래 ‘중학교를 졸업한 후에 어떤 고등학교에 진학할 계획입니까?’에 대한 9가지의 응답 중 ‘특수목적고1(외국어고, 과학고, 국제고)’를 응답한 학생들에게만 값 1을 부여하고 나머지 응답에 대해서는 0을 부여하여 산출하였다. 고등학교 분류체계상 같은 특수목적고로 분류되는 예술고나 체육고, 마이스터고의 경우는 제외하였는데 이는 해당 고등학교는 입시 위주의 외국어고, 과학고, 국제고와 달리 매우 높은 수준의 학업성취도를 요구하지 않기 때문이다. 이어서 부모 교육연한 변인의 경우, 최종학력 대신 각 학력에 해당하는 교육연한을 산출한 후 남자 보호자와 여자 보호자의 값을 평균하여 구했다.

다음으로 경향점수 산출에서 사용하는 9개의 공변인들은 처치가 일어나기 이전 시점인 5차년도 자료의 값을 사용하였다. 여기서도 마찬가지로 선

행연구를 토대로 사교육 참여와 관련이 있다고 밝혀진 변인들을 선정하였으며 이 중 수업태도, 교사에 대한 평가, 혼자 공부시간 변인은 위에서 언급한 방법과 동일하게 산출하였다. 또한 보호자 학습지원과 자기통제력, 자녀학교 만족도 변인도 5점척도의 리커트 척도를 가지며 각 구인에 해당하는 세부 문항 응답을 평균 내어 사용하였다. 다만 월 평균 가정소득의 경우에 정적 편포로 인하여 단변량 정규성을 만족하지 않아 1을 더한 뒤 자연로그 변환을 하여 로그 월소득 변인을 산출하였다.

마지막으로 공변인에 대한 결측처리는 연속변인의 경우 SPSS에서 EM대체 하였으며 범주변인은 STATA에서 다중대체를 1회 실시한 값으로 투입하였다.⁸⁾

8) 모든 공변인에 대하여 결측 비율을 분석한 결과, 10%이하인 것으로 나타났으며 Little의 MCAR 검정을 실시한 결과, 완전임의결측이 아닌 것으로 드러나 결측값에 대한 대체를 시행하였다.

<표 III-2> 투입된 변인 설명

구분	변인	문항내용
종속 변인	학업성취도	6차년도 교과별 수직척도화 점수
처치 변인	사교육 참여	6차년도 기준 지난 1년 동안 교과별 사교육 참여 여부 미참여(0), 참여(1)
공 변 인 1 9)	수업태도	교과별 5점척도 문항의 평균 1) 나는 학교 수업시간에 집중한다. 2) 나는 학교 수업시간에 적극적으로 참여한다. 3) 나는 학교 숙제를 꼬박꼬박 한다. 4) 나는 학교 수업시간에 배운 것을 복습한다. 5) 나는 학교 수업시간에 배울 내용을 미리 예습한다.
	교사에 대한 평가	교과별 5점척도 문항의 평균 1) 선생님은 수업을 열심히 하신다. 2) 선생님은 담당 교과에 대한 지식이 많으시다. 3) 선생님은 수업내용을 알기 쉽게 잘 가르치신다. 4) 선생님은 학생들이 열심히 공부하기를 원하신다. 5) 선생님은 학생들의 높은 성취 수준을 기대하신다. 6) 선생님은 과제를 꼼꼼하게 검사하신다. 7) 선생님은 학생들이 수업 중에 얼마나 잘 이해하고 있는 지 확인하신다. 8) 선생님께서는 토론과 실험, 실습 등 학생 참여형으로 수 업을 진행하신다.
	혼자 공부시간	일주일 동안 혼자 숙제를 하거나 공부하는 시간 1시간 미만(1), 1시간~2시간(2), 2시간~3시간(3), 3시간~4시 간(4), 4시간~5시간(5), 5시간~6시간(6), 6시간~7시간(7), 7시 간~8시간(8), 8시간 이상(9)
	학습방법	5점척도 문항의 평균 1) 나는 새롭게 배운 내용을 이미 알고 있는 것과 어떻게 연결시킬지 생각해 본다. 2) 나는 중요한 내용들은 요점 정리를 하거나 표나 마인드 맵 등으로 정리하여 공부한다. 3) 수업내용들을 잘 이해하고 있는지 스스로 확인한다.
	특목고 진학희망	특목고(외국어고, 과학고, 국제고)(1), 그 외 응답(0)
	부모 교육연한	초등학교 졸업 이하(1), 중학교 졸업(2), 고등학교 졸업(3), 대학교 졸업(4), 4년제 대학 졸업(5), 대학원 석사(6), 대학 원 박사(7)를 교육연한 변환 후 평균
	사립학교	공립(0), 사립(1)
	여학생	남학생(0), 여학생(1)

9) 공변인1은 분위회귀분석모형에 투입된 변인으로서, 이전연도에서만 조사되었던 부모 교육연한, 사립학교, 여학생여부를 제외한 변인들은 6차년도 조사 결과의 값이

공 변 인 2	수업태도	위와 동일
	교사에 대한 평가	위와 동일
	혼자 공부시간	위와 동일
	보호자 학습지원	보호자에 대한 5점척도 문항평균 1) 내가 공부하는 것을 도와주신다. 2) 내가 숙제를 하는지 안 하는지 확인하신다. 3) 학교생활에 대해서 관심을 가지고 물어보신다. 4) 학교 공부를 열심히 하도록 격려해주시다. 5) 나와 많은 시간을 함께 보내려고 노력하신다. 6) 나에게 사랑과 애정을 보이신다. 7) 나를 잘 이해해 주신다. 8) 나와 서로의 일상에 대한 대화를 많이 하신다. 9) 나와 진로에 대한 대화를 많이 하신다. 10) 나와 친구관계에 대한 대화를 많이 하신다. 11) 나와 학교성적에 대한 대화를 많이 하신다. 12) 나와 뉴스나 신문기사와 관련한 대화를 많이 하신다. 13) 내가 어려운 일이 생겼을 때 가장 먼저 상의하는 분이 다.
	자기통제력	5점척도 문항평균 1) 나는 누가 지켜보지 않아도 정해진 규칙이나 지시를 잘 따른다. 2) 나는 주어진 과제가 어려워도 쉽게 포기하지 않는다. 3) 나는 일을 하기 전에 항상 생각을 먼저하고 행동한다.
	로그 월소득	가정의 월평균 가구소득(상여금, 재산소득, 생활보조금 등 도 포함, 만원 단위)에 1을 더한 후 자연로그를 취함
	자녀학교 만족도	5점척도 문항평균 1) 학교는 자녀의 학습 능력을 잘 길러주고 있다. 2) 학교는 자녀의 특기와 적성을 잘 계발해 주고 있다. 3) 학교 선생님들은 학생들을 가르치는데 열의가 있다. 4) 학교에서는 자녀의 수준에 맞게 잘 가르치고 있다. 5) 학교는 자녀에 대한 개별상담과 진로지도를 잘 하고 있 다. 6) 학교는 학습에 불편함이 없는 시설과 환경을 갖추고 있 다. 7) 학교는 자녀의 안전을 위해 노력하고 있다.
	사전 사교육 참여	5차년도 기준 지난 1년 동안 교과별 사교육 참여 여부 미참여(0), 참여(1)
	사전 성취도	5차년도 교과별 수직적도화 점수

며, 공변인2는 경향점수 산출시에 투입된 변인으로 5차년도 조사 결과를 이용하였
다.

3. 연구모형

가. 경향점수 역확률 가중치

경향점수는 처치집단에 속할 확률을 의미하는데, 이 연구에서 경향점수란 학생 개인이 사교육 참여집단에 속할 확률을 말한다. 이 연구의 처치변인은 사교육 참여여부에 대한 이분변인이므로 로지스틱 회귀분석(logistic regression) 또는 프로빗 회귀분석(probit regression)을 실시하여 경향점수를 산출할 수 있다(김준엽, 정혜경, Michael, 2008). 따라서 이 연구에서는 경향점수 산출을 위해 로지스틱 회귀분석을 실시하였으며 이에 해당하는 식은 아래와 같다. 여기서 X 는 5차년도 공변인을, P_i 는 각 개별 관측치의 경향점수를 의미한다.

$$\ln \left[\frac{P(T=1|X)}{P(T=0|X)} \right] = \beta X \quad , \quad P_i = \frac{e^{\beta X}}{1 + e^{\beta X}} = \frac{1}{1 + e^{-\beta X}} \quad \langle \text{식 3.25} \rangle$$

산출한 경향점수를 활용하는 방법에는 크게 매칭(matching)방법, 층화(stratification)방법, 역확률가중치(Inverse Probability Treatment Weight; IPTW) 방법이 있으나 매칭과 층화의 경우 분위회귀분석을 위해 활용하기에 적절하지 않아 Firpo(2007)가 사용하였던 역확률 가중치 방법을 활용하며, <식 3.26>의 방법을 이용하여 각 관측치에 대한 가중치를 산출한다.

$$\hat{w}_i = \frac{T_i}{\hat{P}_i} + \frac{(1 - T_i)}{1 - \hat{P}_i} \quad \langle \text{식 3.26} \rangle$$

그러나 경향점수가 0이나 1에 매우 근접하면서 경향점수와 실제 처치와는 상반될 경우, 해당 가중치는 그렇지 않은 경우들에 비해 그 값이 지나치게 커질 수 있다. 그리고 이는 곧 소수의 관측치가 전체 표본을 과중하게 대표하게 만들면서 추정 결과의 외적 타당성을 확보하기가 어려워진다.

이러한 점을 보완하기 위해 이 연구에서는 안정화 역확률 가중치 (stabilized inverse probability weight)를 사용하였으며 이를 산출하기 위한 계산 방법은 아래와 같다.

$$sw_i = \frac{T_i}{P_i} \cdot \frac{n_t}{\sum_{k=1}^{n_t} w_k} + \frac{(1 - T_i)}{1 - P_i} \cdot \frac{n_c}{\sum_{k=1}^{n_c} w_k} \quad \langle \text{식 3.27} \rangle$$

나. 분위회귀분석모형

분위회귀분석 결과에 대하여 선택편의 통제가 미치는 영향을 알아보기 위해 역확률 가중치를 적용하기 전 분위회귀분석과 역확률 가중치를 적용하여 선택편의를 통제한 분위회귀분석을 각각 실시하였다. 또한 조건부 분위회귀분석과 무조건부 분위회귀분석의 의미가 다르다는 것에서부터, 두 분위회귀분석 각각에 대해 선택편의 통제의 결과를 살펴보았다.

조건부 분위회귀분석은 이론적 배경에서도 이미 살펴보았듯, 효과를 추정하기 원하는 독립변인과 공변인을 함께 투입하는 분위회귀분석이다. 이때, 분위회귀분석의 종속변인은 6차년도 학업성취도이며 처치변인은 6차년도 사교육 참여 여부에 해당한다. 여기서 공변인은 선행연구에 의해 학업성취도에 영향을 미치는 것으로 나타난 변인을 투입하되, 6차년도에 수집된 자료로 얻은 데이터를 활용하였다. 또한 무조건부 분위회귀분석에서는 다른 공변인을 투입하지 않고, 6차년도 사교육 참여 여부만 투입하여 분위회귀분석을 실시하였다.

이 연구에서는 최하위권, 하위권, 중위권, 상위권, 최상위권에 대한 사교육 효과를 밝혀내고자 0.1, 0.25, 0.5, 0.7, 0.9까지 다섯 개의 분위에 대한 회귀분석을 실시하였다. 위의 다섯 개의 분위기를 선정한 것은 첫째로, 상위권과 중위권 하위권 학생들의 사교육 효과가 다르게 나타난다는 기존 연구(오영수, 윤정식, 2003; 차동춘, 2015)를 반영함과 동시에, 성적격차가 본격적으로 벌어지는 중학교 시기의 최상위권과 최하위권에 대한 사교육 효

과 분석을 추가적으로 진행하기 위함이다. 둘째로, 기존에 분위회귀분석을 이용해 사교육 효과를 분석하였던 선행연구(김민선, 백일우, 2016; 이광현, 권용재, 2011)의 결과와 대조하고, 최하위권, 하위권, 중위권, 상위권, 최상위권 중에서 어떤 분위에 대한 선택편의가 가장 두드러지게 일어나는지 분석하기 위함이다.

다. 선택편의가 통제된 분위회귀분석모형

선택편의가 통제된 분위회귀분석을 위해, 조건부 분위회귀분석과 무조건부 분위회귀분석을 같은 방법으로 실시하되, 다만 여기에서는 이전에 산출한 안정화 역확률 가중치를 `pweight`로 투입하였다.

기술통계와 경향점수 산출, 선택편의를 통제하기 전 분위회귀분석과 통제 한 후 분위회귀분석 모두 STATA 14.0에서 진행하였으며 경향점수 산출을 위해서는 `logit`과 `predict` 코드, 분위회귀분석을 위해서는 `qreg` 코드를 이용하였다.

한편, 선택편의를 통제한 무조건부 분포의 사교육 효과를 구할 때는 Frolich와 Melly(2008)의 `ivqte`코드를 이용할 수 있지만 이 경우 안정화 역확률 가중치 대신 기존의 역확률 가중치를 사용한다는 한계가 있어, 해당 분석 역시 마찬가지로 `qreg` 코드를 활용하였다.¹⁰⁾

10) 추가적으로 `ivqte`코드를 활용한 경우와 `qreg`에 `pweight`을 적용하여 분석한 결과를 비교해 보았는데 두 분석 결과의 차이가 거의 없었으며 있는 경우라도 특정 분위에서 계수 값의 차이가 미미하게 났다. 이는 `ivqte`에서는 일반 역확률 가중치를 사용하는 반면, 연구자는 안정화 역확률 가중치를 사용하였기 때문인 것으로 보인다.

IV. 연구결과

1. 기술통계

주요변인의 기술통계치를 제시하면 다음과 같다.

<표 IV-1> 수학교과 주요변인 기술통계(N=3,289)

구분	주요변인	평균	표준 편차	최솟값	최댓값	왜도	첨도
6차년도	수학 성취도	357.22	63.71	176.00	515.00	0.51	-0.46
	사교육 참여	0.75	0.43	0.00	1.00	-1.14	-0.70
	수업태도	3.82	0.93	1.00	5.00	-0.61	0.06
	교사에 대한 평가	4.19	0.72	1.00	5.00	-0.75	0.60
	혼자 학습시간	2.86	1.94	0.65	9.00	1.47	1.93
	학습방법	3.33	0.88	1.00	5.00	-0.14	0.15
	특목고 진학희망	0.06	0.23	0.00	1.00	3.83	12.67
1차년도	부모 교육연한	14.51	2.10	6.00	22.00	-0.02	0.13
4차년도	사립학교	0.25	0.43	0.00	1.00	1.18	-0.61
	여학생	0.48	0.50	0.00	1.00	0.08	-1.99
5차년도	수업태도	3.76	0.94	1.00	5.00	-0.48	-0.21
	교사에 대한 평가	4.14	0.77	1.00	5.00	-0.85	0.88
	혼자 학습시간	3.03	2.02	0.34	9.00	1.30	1.29
	보호자 학습지원	3.86	0.75	1.00	5.00	-0.38	-0.08
	자기통제력	3.57	0.75	1.00	5.00	-0.04	0.30
	로그 월소득	6.06	0.62	0.00	9.05	-1.49	14.57
	자녀학교 만족도	3.54	0.68	1.00	5.00	-0.28	0.68
	사교육 참여	0.75	0.43	0.00	1.00	-1.14	-0.71
	수학 성취도	339.77	58.84	162.00	505.00	0.47	-0.16

<표 IV-2> 영어교과 주요변인 기술통계(N=3,251)

구분	주요변인	평균	표준 편차	최솟값	최댓값	왜도	첨도
6차년도	영어 성취도	399.30	67.87	240.00	519.00	0.29	-1.01
	사교육 참여	0.71	0.45	0.00	1.00	-0.92	-1.15
	수업태도	3.80	0.90	1.00	5.00	-0.51	-0.03
	교사에 대한 평가	4.22	0.71	1.00	5.00	-0.75	0.51
	혼자 학습시간	2.63	1.80	0.97	9.00	1.58	2.64
	학습방법	3.33	0.88	1.00	5.00	-0.14	0.16
	특목고 진학희망	0.06	0.23	0.00	1.00	3.80	12.47
1차년도	부모 교육연한	14.52	2.12	6.00	22.00	0.00	0.16
4차년도	사립학교	0.25	0.43	0.00	1.00	1.18	-0.61
	여학생	0.48	0.50	0.00	1.00	0.07	-2.00
5차년도	수업태도	3.76	0.91	1.00	5.00	-0.38	-0.30
	교사에 대한 평가	4.18	0.74	1.00	5.00	-0.80	0.74
	혼자 학습시간	2.81	1.92	1.00	9.00	1.41	1.73
	보호자 학습지원	3.86	0.76	1.00	5.00	-0.37	-0.10
	자기통제력	3.57	0.75	1.00	5.00	-0.03	0.31
	로그 월소득	6.06	0.62	0.00	9.05	-1.49	14.63
	자녀학교 만족도	3.53	0.68	1.00	5.00	-0.30	0.70
	사교육 참여	0.73	0.45	0.00	1.00	-1.03	-0.95
	영어 성취도	392.75	68.67	180.00	507.00	0.19	-0.90

<표 IV-2>에서 제시한 바와 같이 사용된 변인의 왜도의 범위는 -1.49에서 3.83으로, 첨도의 범위는 -2에서 14.57로 산출되어 대체로 단변량 정규성을 크게 위배하지 않는 것으로 나타났다. 또한 각 교과별 사교육 참여(참여=1, 미참여=0)에 대한 평균은 수학은 0.75, 영어는 0.71로 나타나 수학교과 사교육에 참여하는 학생의 비율은 75%, 영어교과 사교육에 참여하는 학생의 비율은 71%임을 알 수 있다.

2. 경향점수 및 안정화 역확률 가중치 추정

<표 IV-3>와 <표 IV-4>는 경향점수 산출을 위해 각 교과별로 로지스틱 회귀분석을 실시한 결과이다. 여기서 종속변수인 사교육 참여 변인은 이분형 변인이기 때문에 선형확률모형(Linear Probability Model; LPM)을 사용할 수도 있겠으나, 이 모형을 사용하는 경우 추정된 경향점수의 값이 0이하 또는 1초과로 산출되는 문제가 발생한다.¹¹⁾ 이러한 이유로 이분형 변인에 대한 경향점수 산출 시에는 대부분 로지스틱 회귀분석(Logistic Regression)이나 프로빗 회귀분석(Probit Regression)을 사용한다. 이 연구에서는 로지스틱 회귀분석을 사용하였다.¹²⁾

<표 IV-3> 수학교과 로지스틱 회귀분석 결과 (N=3,289)

구분	계수	표준오차	z	유의확률
수학 사교육 참여				
수업태도	0.185	0.068	2.720	0.006
교사평가	-0.088	0.073	-1.200	0.231
혼자 학습시간	0.029	0.030	0.970	0.332
학습방법	0.058	0.073	0.790	0.428
자기통제력	-0.025	0.078	-0.330	0.743
로그 월소득	0.547	0.094	5.810	0.000
자녀학교 만족도	-0.037	0.071	-0.520	0.603
이전 사교육참여	1.980	0.102	19.500	0.000
학업성취도	0.007	0.001	6.810	0.000
부모 교육연한	0.062	0.027	2.270	0.023

11) 이 외에도 선형 확률 모형은 한계 효과(Marginal Effect)가 일정하다고 가정한다는 점, 이분산성을 가진다는 점 등이 경향점수 분석에 적절치 않다.

12) 처치변인이 이분형이 아닌 경우, 로지스틱 분석과 프로빗 분석의 경향점수 산출 결과는 다소 상이할 수 있으나 이분형의 경우, 두 분석의 결과는 유사하다(Caliendo & Kopeinig, 2005).

<표 IV-4> 영어교과 로지스틱 회귀분석 결과 (N=3,251)

구분	계수	표준오차	z	유의확률
영어 사교육 참여				
수업태도	0.120	0.064	1.860	0.063
교사평가	0.049	0.072	0.690	0.493
혼자 학습시간	0.075	0.029	2.590	0.009
학습방법	-0.012	0.069	-0.170	0.863
자기통제력	0.065	0.073	0.890	0.371
로그 월소득	0.492	0.089	5.530	0.000
자녀학교 만족도	-0.030	0.068	-0.450	0.655
이전 사교육참여	1.988	0.097	20.590	0.000
학업성취도	0.004	0.001	5.460	0.000
부모 교육연한	-0.003	0.025	-0.110	0.910

그 결과 수학교과와 영어교과 모두의 경우, 수업태도와 로그 월소득, 이전 사교육참여, 이전연도 학업성취도, 부모 교육연한이 사교육 참여에 유의한 영향을 미치는 것으로 나타났다. 또한 영어교과와 영어교과 영어교과 모두의 경우 혼자학습시간, 로그 월소득, 이전 사교육참여, 이전연도 학업성취도가 사교육 참여에 유의한 영향을 미치는 것으로 나타났다. 이를 좀 더 자세히 살펴보면 로그 월소득과 이전 사교육 참여, 학업성취도(이전연도)는 수학과 영어 교과 모두에서 사교육 참여에 유의한 영향을 미쳤던 반면 수업태도와 부모교육연한, 혼자학습시간은 각각 수학과 영어 교과에만 유의하다는 것을 알 수 있다.

또한 유의한 모든 변인들은 각 교과별 사교육 참여에 정적으로 유의한 것으로 나타나, 수업태도가 좋을수록, 로그 월소득이 높을수록, 이전연도 사교육에 참여할수록, 혼자 학습시간이 많을수록, 이전연도 학업성취도와 부모 교육연한이 높은 학생일수록 사교육에 참여할 확률이 높은 것으로 나타났다.

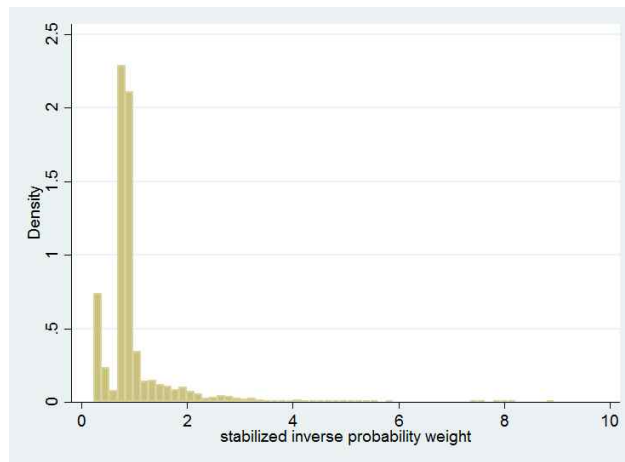
한편, 로지스틱 회귀분석을 통해 얻은 각 교과별 경향점수는 predict 명령어를 통해 저장한 후 역확률 가중치와 안정화 역확률 가중치를 각각 산출하였다.

각 교과별로 산출된 역확률 가중치와 안정화 역확률 가중치에 대한 결과는 <표 IV-5>와 같다. 역확률 가중치의 경우 최댓값이 수학과 영어 각각에서 35.405와 24.529로 산출됨에 따라, 하나의 관측치에 과도하게 높은 가중치가 부여될 위험이 있었던 데 비해, 안정화 역확률 가중치는 그 값이 8.948과 7.191로 비교적 안정적이라는 것을 알 수 있다. 따라서 비교적 안정적인 안정화 역확률 가중치를 사용하였으며, 이 때, 구해진 가중치의 최댓값이 10을 넘지 않았기 때문에 추가적인 절삭(truncation)은 실시하지 않았다. 산출된 안정화 역확률 가중치의 분포는 각 교과별로 <그림 IV-1>과 <그림 IV-2>와 같다.

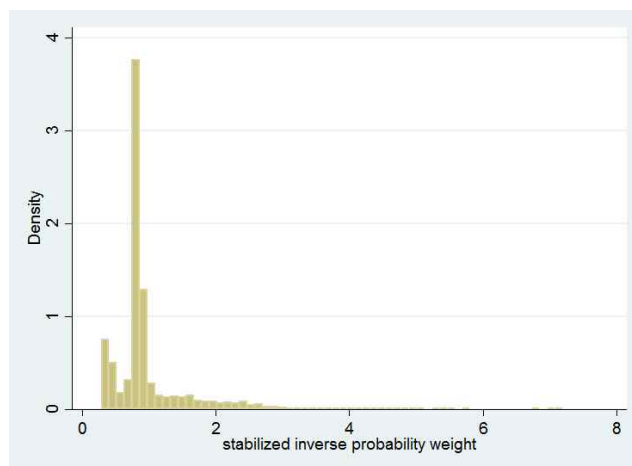
한편, 산출된 안정화 역확률 가중치가 사전 공변인의 영향을 통제할 수 있는지 알아보기 위해 로지스틱 회귀분석에 투입된 각 공변인에 대하여 가중 t검정을 실시하였다. 그 결과 모든 사전 공변인에 대한 균형(balance)이 확보된 것을 <표 IV-6>과 <표 IV-7>을 통해 확인할 수 있다.

<표 IV-5> 역확률 가중치 산출결과

교과	구분	평균	표준편차	최솟값	최댓값
수학	역확률가중치	1.965	2.470	1.011	35.405
	안정화 역확률가중치	0.994	0.703	0.256	8.948
영어	역확률가중치	1.995	2.054	1.019	24.529
	안정화 역확률가중치	0.999	0.664	0.296	7.191



<그림 IV-1> 수학교과 안정화 역확률 가중치



<그림 IV-2> 영어교과 안정화 역확률 가중치

<표 IV-6> 경향점수 반영 후 수학교과 밸런스 테스트 결과 (N=3,289)

구분	계수	표준오차	t	유의확률
수학 사교육 참여				
수업태도	-0.006	0.011	-0.500	0.616
교사평가	0.004	0.015	0.290	0.775
혼자 학습시간	0.002	0.006	0.320	0.746
학습방법	-0.001	0.014	-0.060	0.955
자기통제력	-0.008	0.014	-0.540	0.587
로그 월소득	-0.002	0.019	-0.100	0.921
자녀학교 만족도	-0.006	0.015	-0.410	0.682
이전 사교육참여	0.016	0.020	0.820	0.414
학업성취도	0.000	0.000	0.760	0.445
부모 교육연한	0.002	0.005	0.300	0.762

<표 IV-7> 경향점수 반영 후 영어교과 밸런스 테스트 결과 (N=3,251)

구분	계수	표준오차	t	유의확률
영어 사교육 참여				
수업태도	-0.007	0.012	-0.560	0.574
교사평가	0.004	0.015	0.300	0.764
혼자 학습시간	-0.005	0.007	-0.790	0.427
학습방법	0.001	0.015	0.070	0.940
자기통제력	-0.014	0.015	-0.910	0.365
로그 월소득	-0.013	0.019	-0.690	0.490
자녀학교 만족도	-0.009	0.016	-0.570	0.569
이전 사교육참여	0.005	0.020	0.250	0.801
학업성취도	0.000	0.000	0.700	0.486
부모 교육연한	-0.001	0.005	-0.170	0.864

3. 분위회귀분석

가. 수학교과에 대한 분석 결과

분위회귀분석에서는 사교육 참여여부와 6차년도 공변인을 함께 투입하는 조건부 분위회귀분석과, 공변인을 투입하지 않은 무조건부 분위회귀분석을 각 교과별로 실시하였다. 또한 이 결과를 단순회귀분석의 결과와 비교하기 위해 모든 조건부 분위회귀분석의 결과표 첫 번째 열에 해당 결과를 삽입하였다.

<표 IV-8> 수학교과 조건부 분위회귀분석 결과 (N=3,289)

구분	OLS	0.1분위	0.25분위	0.5분위	0.75분위	0.9분위
6차 학업성취도						
사교육 여부	28.908*** (2.274)	12.326*** (1.913)	15.955*** (1.854)	26.342*** (2.095)	38.768*** (2.702)	47.188*** (3.945)
수학수업 자세	11.767*** (1.351)	5.671*** (1.118)	8.302*** (1.078)	12.574*** (1.125)	13.562*** (1.820)	16.626*** (2.202)
수학교사 평가	3.071* (1.475)	0.161 (0.872)	2.737* (1.132)	3.341* (1.302)	4.105* (1.895)	6.860** (2.272)
혼자학습 시간	4.998*** (0.539)	2.354** (0.804)	4.180*** (0.705)	5.921*** (0.694)	6.964*** (0.785)	6.222*** (0.877)
학습 방법	5.914*** (1.261)	0.952 (1.128)	2.941** (1.116)	5.969*** (1.242)	7.705*** (1.674)	8.578*** (2.102)
특목고 진학희망	30.864*** (4.145)	27.091* (11.807)	44.750*** (6.686)	34.346*** (3.686)	33.135*** (7.204)	19.003** (6.747)
부모평균 교육연한	4.903*** (0.477)	1.219* (0.473)	2.574*** (0.442)	5.061*** (0.470)	6.028*** (0.657)	6.370*** (0.817)
사립여부	0.950 (2.160)	2.209 (2.213)	1.308 (2.232)	0.181 (2.228)	-0.737 (2.915)	2.176 (3.842)
여학생	8.078*** (1.863)	10.005*** (1.779)	11.005*** (1.730)	7.607*** (1.890)	3.297 (2.557)	-2.596 (3.213)
절편	166.779*** (8.487)	227.988*** (6.526)	196.015*** (7.003)	155.899*** (7.776)	156.49*** (11.311)	160.249*** (13.647)
R^2	0.3027					
$pseudo R^2$		0.0581	0.1193	0.2134	0.2323	0.2003

***p<.001, **p<.01, *p<.05

수학교과에 대한 분석 결과를 살펴보면 다음과 같다. 먼저, 단순회귀분석 결과의 설명량(R^2)과 조건부 분위회귀분석의 설명량($pseudo R^2$)을 보면, $pseudo R^2$ 의 값이 R^2 와 비교했을 때 비교적 낮은 값을 가지는데, 사실이 두 지표의 정의는 서로 다르므로 직접적인 비교는 어렵다.¹³⁾ 다만 조건부 분위회귀분석 내에서 보았을 때, 하위분위 보다 중상위분위에서 상대적으로 높은 모형의 적합도를 보여준다는 것은 확인할 수 있다.

이어서 공변인을 통제한 조건부 분위회귀분석에서, 사교육 참여가 학업성취도에 미치는 영향을 살펴보면 다음과 같다. 모든 분위에서 사교육 참여는 학업성취도에 유의한 영향을 미쳤지만 그 효과의 크기는 상이하게 나타났다. 먼저 0.1분위에서는 사교육 참여의 영향이 12.326으로 나타났지만 분위가 증가할수록 큰 폭으로 증가하여 0.9분위에서는 47.188으로 나타났다. 이는 사교육 참여가, 학업성취도에 영향을 미칠만한 6차년도 공변인을 통제하였을 때 성적 분위가 높을수록 영향력이 증가한다고 볼 수 있다.¹⁴⁾ 이는 김민선과 백일우(2016)의 연구에서 수학 사교육 비용의 효과가 높은 분위로 갈수록 유의해지며, 그 크기도 증가했던 결과와 어느 정도 일관된다고 할 수 있다.¹⁵⁾ 한편, 단순회귀분석에서도 분위회귀분석의 결과와 마찬가지로 사교육 참여가 학업성취도에 미치는 영향이 유의하였고 계수

13) 단순회귀분석에서는 오차의 등분산성을 가정하고 있었던데 반해 분위회귀분석에서는 오차의 이분산성을 가정하고 있다. 따라서 분위회귀분석에서는 단순회귀분석에서 사용하는 R^2 대신 $pseudo-R^2$ 를 이용하여 분위별 모형적합도를 보여준다.

14) 이 연구에서 최종적으로 보고자 하는 것은 선택편의를 통제하기 전과 후의 각 분위별 계수와 유의도가 어떻게 달라지는가에 있지, 분위에 따른 사교육 효과크기를 비교하는 것은 주요 관심사가 아니다. 따라서 계수의 크기와 관련하여, 상대적으로 낮은 적합도를 가졌던 0.1분위에서의 계수와 높은 적합도를 가졌던 0.9분위와 비교하는 것은, 각 분위에 따라 효과 크기가 다른 것이 아니라 적합도 차이로부터 기인했을 가능성이 있으므로 특별히 해석에 유의하여야 한다.

15) 그러나 통제 변인의 투입에 따라 분포와 분위수가 달라질 수 있는 조건부 분위회귀분석의 특성상 이 연구를 김민선과 백일우(2016)의 연구와 완전히 동일선상에서 비교할 수는 없다.

의 크기가 0.1분위보다는 높았고 0.9분위보다는 작았으며 중간값에 해당하는 0.5분위의 계수와 유사한 값을 보였다.

<표 IV-9> 수학교과 무조건부 회귀분석 결과 (N=3,289)

분위	Δ	표준오차	t	유의확률	95% 신뢰구간	
0.1	17	3.684	4.610	0.000	9.777	24.223
0.25	26	2.829	9.190	0.000	20.453	31.547
0.5	59	3.605	16.370	0.000	51.932	66.068
0.75	76	4.177	18.190	0.000	67.809	84.191
0.9	87	7.511	11.580	0.000	72.274	101.726

한편, <표 IV-9>에서는 학업성취도에 영향을 미칠만한 6차년도 공변인을 투입하지 않은 무조건부 분위회귀 분석의 결과를 보여주고 있다. 이 표를 통해, 사교육 참여 집단 내 τ 분위에 해당하는 학업성취도와 사교육 미참여 집단 내에 있는 τ 분위에 해당하는 학업성취도 차이를 알 수 있다. 좀 더 자세히 살펴보면 모든 분위에서 사교육 참여가 학업성취도에 유의한 영향을 미치는 것을 확인할 수 있으며 차이의 크기 또한 분위가 증가할수록 높아진다.

나. 영어교과에 대한 분석 결과

영어교과에서도 마찬가지로 사교육 참여가 학업성취도에 미치는 영향은 단순회귀분석과 조건부 분위회귀분석의 전 분위에서 모두 정적으로 유의하게 나타났다. 즉, 모든 성적분위에서 사교육을 참여하는 학생의 학업성취도가 높다는 것을 확인할 수 있다. 그러나 최하위 분위인 0.1분위에서 상위 분위인 0.75분위까지는 분위가 증가할수록 사교육 참여 효과의 크기가 증가하였던 데 반해, 최상위권인 0.9분위에서는 그 크기가 다시 감소하는 것으로 나타났다.

<표 IV-10> 영어교과 조건부 분위회귀분석 결과 (N=3,251)

구분	OLS	0.1분위	0.25분위	0.5분위	0.75분위	0.9분위
6차 학업성취도						
사교육 여부	22.622*** (2.306)	10.108*** (2.515)	17.109*** (1.795)	25.971*** (2.642)	27.048*** (3.045)	25.475*** (4.325)
수학수업 자세	8.579*** (1.498)	4.180** (1.575)	7.203*** (1.057)	10.049*** (1.742)	11.704*** (2.019)	9.433*** (2.650)
수학교사 평가	3.421* (1.627)	2.242 (1.643)	3.332** (1.059)	3.608 (1.844)	3.546 (2.155)	3.259 (2.741)
혼자학습 시간	3.990*** (0.613)	2.270** (0.734)	3.514*** (0.847)	4.507*** (0.740)	5.078*** (0.781)	4.006*** (0.659)
학습 방법	8.941*** (1.357)	2.778 (1.595)	4.271*** (1.076)	11.466*** (1.659)	11.686*** (1.758)	12.638*** (2.111)
특목고 진학희망	39.625*** (4.416)	31.621** (11.661)	60.980*** (8.704)	48.447*** (4.597)	34.103*** (5.431)	25.600*** (4.110)
부모평균 교육연한	7.438*** (0.500)	3.169*** (0.586)	5.398*** (0.467)	8.368*** (0.599)	9.169*** (0.673)	8.355*** (0.763)
사립여부	3.432 (2.318)	-5.684* (2.869)	-0.263 (2.739)	4.919 (2.974)	4.844 (3.312)	1.196 (3.788)
여학생	22.215*** (1.996)	16.891*** (2.336)	23.546*** (1.973)	21.993*** (2.476)	20.037*** (2.776)	20.993*** (3.178)
절편	174.114*** (9.062)	224.707*** (9.798)	186.486*** (7.401)	139.599*** (10.04)	162.754*** (12.492)	220.608*** (15.272)
R^2	0.303					
$pseudo R^2$		0.0698	0.1360	0.2106	0.2160	0.1711

***p<.001, **p<.01, *p<.05

한편, <표 IV-11>에서는 공변인을 투입하지 않은 무조건 분위회귀분석 결과를 살펴볼 수 있는데 0.1분위에서 0.5분위까지는 분위가 증가할수록 사교육 참여집단과 미참여집단의 학업성취도 차이가 점점 증가하였으나 0.75분위부터는 다시 감소하는 것을 확인할 수 있다. 0.9분위부터 계수의 값이 감소하였던 조건부 분위회귀분석의 결과와는 다소 다른 결과인데 이는 앞서 이론적 배경에서 설명하였던 바와 같이 조건부 분위회귀분석에서의 τ 분위와 무조건부 분위회귀분석에서의 τ 분위는 다르기 때문이다.

<표 IV-11 > 영어교과 무조건부 회귀분석 결과 (N=3,251)

분위	Δ	표준오차	t	유의확률	95% 신뢰구간	
0.1	11	2.439	4.510	0.000	6.218	15.782
0.25	31	2.677	11.580	0.000	25.751	36.249
0.5	58	3.508	16.530	0.000	51.121	64.879
0.75	47	4.738	9.920	0.000	37.709	56.291
0.9	40	1.732	23.100	0.000	36.604	43.396

4. 선택편의가 통제된 분위회귀분석

가. 수학교과에 대한 분석 결과

이 장에서는 사교육 선택에 대한 편의를 통제한 후 각 교과에 대해 조건부 분위회귀분석과 무조건부 분위회귀분석을 실시한 결과를 다룬다. 이는 앞서 이론적 배경에서도 설명하였듯이 분위회귀분석에서도 단순회귀분석과 마찬가지로 사교육 참여에 대한 처치가 무작위로 일어났다는 가정이 성립될 때에만 사교육 참여에 대한 계수를 인과효과로 해석할 수 있기 때문이다. 따라서 바로 이전과 동일한 방식으로 분위회귀분석을 실시하되, 앞서 산출한 안정화 역확률 가중치를 적용하였다.

<표 IV-12> 수학교과 조건부 분위회귀분석 결과 (N=3,289)

구분	OLS	0.1분위	0.25분위	0.5분위	0.75분위	0.9분위
6차 학업성취도						
사교육 여부	22.137*** (2.887)	11.343*** (2.428)	10.169*** (2.256)	16.892*** (2.981)	23.063*** (2.972)	29.641*** (6.317)
수학수업 자세	13.616*** (1.542)	5.694*** (0.970)	7.835*** (1.147)	13.962*** (1.390)	17.932*** (2.129)	20.315*** (2.089)
수학교사 평가	1.630 (1.611)	-1.163 (1.062)	1.935 (1.219)	3.889* (1.694)	2.923 (2.060)	6.937** (2.375)
혼자학습 시간	4.847*** (0.705)	1.809* (0.920)	3.502*** (0.728)	6.275*** (0.937)	6.643*** (0.931)	5.979*** (0.773)

학습 방법	5.031** (1.517)	0.690 (1.210)	2.867* (1.167)	4.656** (1.552)	7.311*** (1.828)	8.336*** (2.239)
특목고 진학희망	31.103*** (4.773)	29.707* (13.547)	49.767*** (7.354)	40.756*** (6.865)	38.069*** (3.911)	26.318*** (3.156)
부모평균 교육연한	5.036*** (0.550)	1.694** (0.538)	3.233*** (0.414)	5.556*** (0.580)	7.228*** (0.726)	7.884*** (0.758)
사립여부	-0.740 (2.389)	1.426 (2.550)	-1.118 (2.027)	-3.406 (2.805)	-1.321 (3.019)	-0.508 (3.551)
여학생	10.213*** (2.077)	11.533*** (2.187)	12.657*** (1.749)	9.746*** (2.355)	3.207 (2.702)	0.762 (3.553)
절편	169.203*** (8.892)	227.097*** (8.724)	195.879*** (6.670)	149.393*** (9.347)	138.159*** (12.652)	136.524*** (14.197)
R^2	0.2579					
$pseudo R^2$		0.0476	0.1016	0.1858	0.2153	0.1755

***p<.001, **p<.01, *p<.05

<표 IV-13> 수학교과 무조건부 회귀분석 결과 (N=3,289)

분위	Δ	표준오차	t	유의확률	95% 신뢰구간	
0.1	8	3.248	2.460	0.014	1.632	14.368
0.25	21	2.328	9.020	0.000	16.435	25.565
0.5	32	3.913	8.180	0.000	24.327	39.673
0.75	35	7.505	4.660	0.000	20.285	49.715
0.9	16	10.220	1.570	0.118	-4.039	36.039

이에 따른 수학교과에 대한 분석 결과는 다음과 같다. 먼저, <표 IV-14>를 통해 알 수 있듯, 공변인을 통제한 조건부 분포에서는 사교육 참여에 대한 선택편의를 통제하였음에도 불구하고, 여전히 모든 분위에서 사교육 참여의 영향이 정적으로 유의한 것으로 나타났다. 그러나 모든 분위에서 계수의 크기가 줄어들었고 분위가 커질수록 선택편의 통제에 따른 계수의 차이가 증가하는 것을 확인할 수 있다. 이는 선택편의를 통제하지 않은 상태에서 조건부 분위회귀분석을 실시하여 사교육 참여의 영향을 추정하는 경우, 실제 영향보다 과대추정(overestimation) 될 가능성을 암시한다. 또한 이 과대추정의 정도는 공변인 통제 후 성적 분위가 높을수록 증가한다.

<표 IV-14> 수학교과 조건부 분위회귀분석 계수 비교

구분	OLS	0.1분위	0.25분위	0.5분위	0.75분위	0.9분위
6차 학업성취도						
통제 전	28.908*** (2.274)	12.326*** (1.913)	15.955*** (1.854)	26.342*** (2.095)	38.768*** (2.702)	47.188*** (3.945)
통제 후	22.137*** (2.887)	11.343*** (2.428)	10.169*** (2.256)	16.892*** (2.981)	23.063*** (2.972)	29.641*** (6.317)
계수 차이	6.771	0.983	5.786	9.450	15.705	17.547

***p<.001, **p<.01, *p<.05

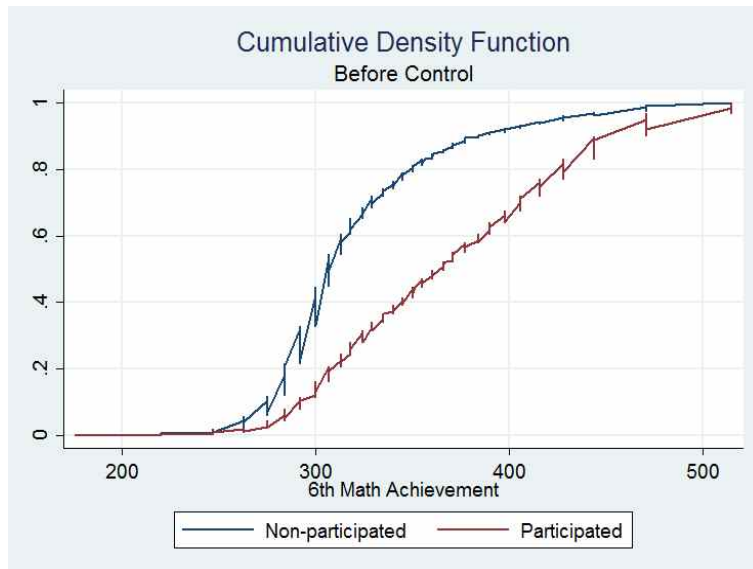
한편, <표 IV-15>에는 선택편의를 통제한 후 무조건부 분위회귀분석을 실시한 결과가 나타나 있다. 선택편의를 통제하기 전 사교육 참여가 학업 성취도에 미치는 영향은 모든 분위에서 .001 수준으로 유의하였던 것에 비해 선택편의를 통제한 후에는 0.1분위의 t값이 감소하여 기존보다 유의확률이 증가하였으며 0.9분위의 경우 더 이상 사교육 참여의 효과가 유의하지 않다는 것을 알 수 있다. 또한 사교육 참여에 대한 계수 값도 현저히 감소하였고, 감소의 폭은 분위가 높아질수록 증가하는 것으로 나타났다. 이는 무조건부 분위회귀분석에서도 마찬가지로, 성적 분위가 높을수록 사교육 효과에 대한 선택편의가 큰 폭으로 발생한다는 것으로, 선택편의를 통제하지 않을 경우, 사교육 참여 효과에 대한 과대추정이 발생할 수 있다는 점을 시사한다.

<표 IV-15> 수학교과 무조건부 분위회귀분석 계수 비교

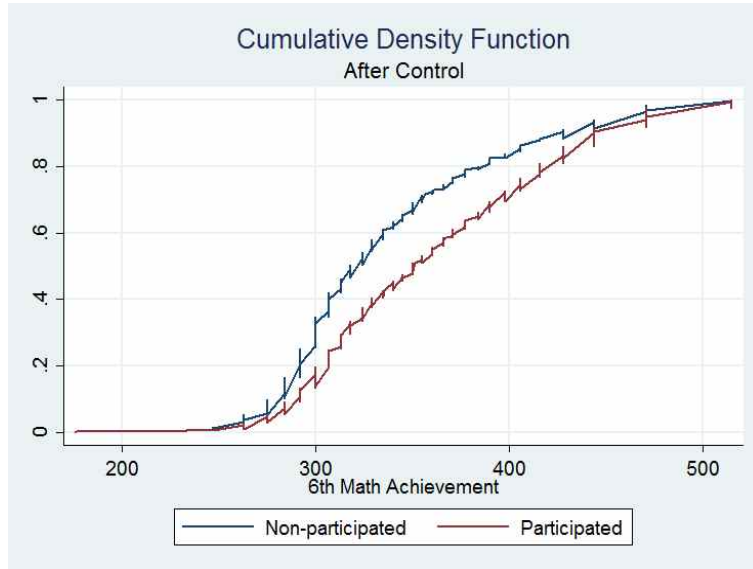
구분	0.1분위	0.25분위	0.5분위	0.75분위	0.9분위
6차 학업성취도					
통제 전	17***	26***	59***	76***	87***
통제 후	8*	21***	32***	35***	16
계수 차이	9	5	27	41	71

***p<.001, **p<.01, *p<.05

무조건부 분위회귀분석의 경우, 선택편의 통제에 따라 사교육 참여 집단과 미참여집단의 누적 확률 분포(Cumulative Distribution Function)가 어떻게 달라지는지 확인할 수 있다. 이에 따라 아래에 제시된 [그림 IV-3]과 [그림 IV-4]은 사교육 참여집단과 미참여 집단의 6차년도 학업성취도에 대한 누적확률분포이다. 이 중 [그림 IV-3]은 사교육 참여에 대한 선택편의를 통제하기 전의 분포를, [그림 IV-4]는 선택편의를 통제하고 나서의 분포를 나타내고 있다. 두 그림을 대조해보면 선택편의를 통제한 후 거의 모든 분위에서 두 확률분포함수 사이의 거리가 좁혀지는 것을 확인할 수 있다. 또한 통제 이후 0.1분위와 0.9분위에서는 두 분포 사이의 거리가 매우 좁아진 것을 확인할 수 있는데, 이는 앞에서 제시한 무조건부 분위회귀분석의 결과가 선택편의 조정에 따라 달라진 것과 같은 맥락에서 해석할 수 있다.



[그림 IV-3] 수학교과 선택편의 통제 전 누적확률분포



[그림 IV-4] 수학교과 선택편의 통제 후 누적확률분포

나. 영어교과에 대한 분석 결과

한편, 영어교과에서도 마찬가지로 안정화 역확률 가중치를 통해 선택편의를 통제한 후 조건부 분위회귀분석과 무조건부 분위회귀분석을 실시하여 분위별 계수가 어떻게 달라지는지 알아보았다.

<표 IV-16> 영어교과 조건부 분위회귀분석 결과 (N=3,251)

구분	OLS	0.1분위	0.25분위	0.5분위	0.75분위	0.9분위
6차 학업성취도						
사교육 여부	7.494* (2.951)	2.085 (2.814)	4.793 (2.896)	10.057** (3.798)	8.040 (4.413)	8.060** (2.553)
수학수업 자세	9.897*** (1.720)	3.235* (1.562)	8.180*** (1.472)	11.141*** (2.028)	12.704*** (2.790)	11.485*** (2.496)
수학교사 평가	2.384 (1.883)	1.972 (1.874)	3.102 (1.661)	2.115 (2.242)	4.865 (2.972)	4.692 (2.525)
혼자학습 시간	3.727*** (0.771)	2.430*** (0.664)	3.304** (0.990)	4.095*** (0.993)	4.732*** (1.116)	4.216*** (0.481)
학습 방법	9.527*** (1.638)	2.090 (1.708)	4.760** (1.410)	11.595*** (1.983)	13.529*** (2.550)	10.895*** (1.697)

특목고 진학희망	43.888*** (5.870)	38.835*** (11.031)	71.134*** (13.272)	52.828*** (4.552)	29.58*** (5.944)	26.502*** (3.520)
부모평균 교육연한	8.429*** (0.589)	3.889*** (0.642)	5.975*** (0.580)	10.083*** (0.735)	11.046*** (0.933)	9.368*** (0.608)
사립여부	1.912 (2.833)	-4.224 (2.912)	-0.214 (3.703)	1.031 (3.317)	2.488 (4.584)	1.883 (3.341)
여학생	23.404*** (2.363)	15.892*** (2.624)	22.573*** (2.312)	23.263*** (3.020)	22.176*** (3.790)	25.377*** (2.607)
절편	167.318*** (10.120)	224.944*** (10.959)	183.478*** (9.373)	127.086*** (12.088)	132.673*** (16.943)	206.957*** (14.286)
R^2	0.2905					
$pseudo R^2$		0.0627	0.1236	0.1992	0.2149	0.1752

***p<.001, **p<.01, *p<.05

<표 IV-17> 영어교과 무조건부 회귀분석 결과 (N=3,251)

분위	Δ	표준오차	t	유의확률	95% 신뢰구간	
0.1	0	2.189	0.000	1.000	-4.291	4.291
0.25	13	3.357	3.870	0.000	6.418	19.582
0.5	23	5.620	4.090	0.000	11.981	34.019
0.75	14	7.205	1.940	0.052	-0.126	28.126
0.9	0	12.963	0.000	1.000	-25.417	25.417

먼저, 공변인을 투입한 조건부 분포에서, 사교육 참여가 가지는 각 분위별 계수는 선택편의 통제 이전에는 모든 분위에서 정적으로 유의하였던 것과 달리 선택편의 통제 이후에는 0.1분위와 0.25분위, 0.75분위에서 더 이상 유의하지 않은 것으로 나타났다. 또한 0.9분위와 0.5분위의 계수는 선택편의 통제 후에도 여전히 유의하게 나타났지만 통제 이전보다 t값이 다소 감소한 것으로 나타났다. 산출된 계수의 크기 또한 큰 폭으로 감소하였는데 수학교과에서는 성적 분위가 높을수록, 선택편의에 따른 사교육 참여의 계수변화의 크기가 증가하였던 반면 영어교과에서는 0.9분위에서 계수의 차이가 다소 감소한다. 이는 영어교과에서는 중상분위에서 사교육 참여에 의한 선택편의가 가장 많이 발생한다는 것을 말한다. 그러나 여전히 모든 분위에서 선택편의 통제 후에 사교육 참여 계수가 감소하는 것으로 볼 때, 수학교과와 마찬가지로 선택편의를 통제하지 않은 채로 무조건부 분위 회귀분석을 실시하는 경우, 각 분위에서 사교육 참여에 대한 효과가 과대

추정 된다는 것을 알 수 있다.

<표 IV-18> 영어교과 조건부 분위회귀분석 계수 비교

구분	0.1분위	0.25분위	0.5분위	0.75분위	0.9분위
6차 학업성취도					
통제 전	10.108*** (2.515)	17.109*** (1.795)	25.971*** (2.642)	27.048*** (3.045)	25.475*** (4.325)
통제 후	2.085 (2.814)	4.793 (2.896)	10.057** (3.798)	8.040 (4.413)	8.060** (2.553)
계수 차이	8.023	12.316	15.914	19.008	17.415

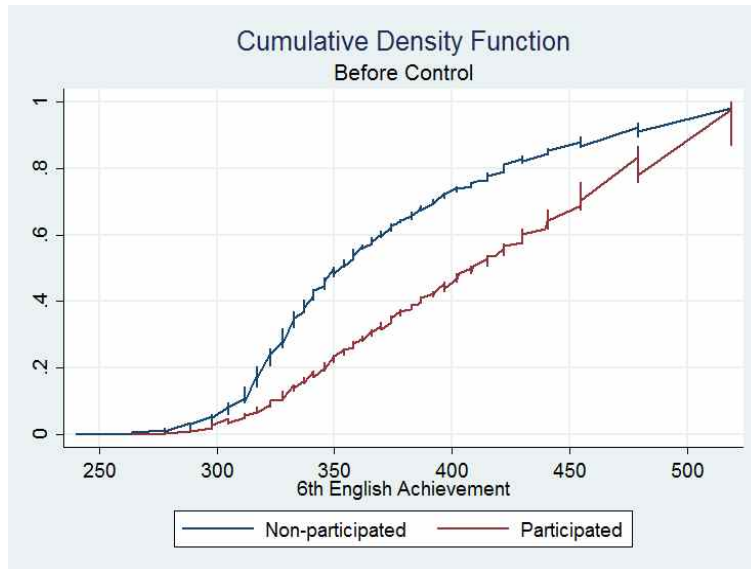
***p<.001, **p<.01, *p<.05

한편, <표 IV-19>에서는 영어 사교육 참여에 대한 선택편의를 통제한 후, 무조건부 분위회귀분석을 실시한 결과가 나타나 있다. 선택편의를 통제하기 전 사교육 참여가 학업성취도에 미치는 영향은 모든 분위에서 .001 수준으로 유의하였던 것에 비해, 선택편의를 통제한 후 최상위 분위와 상위 분위, 최하위 분위에서 경우 더 이상 사교육 참여의 효과가 유의하지 않다는 것을 알 수 있다. 이는 최상위 분위와 최하위 분위에서 나타나는 학업성취도의 차이는 사교육 참여에 의한 효과처럼 보일 수 있으나, 실제로는 선택편의로 인한 것임을 말한다. 동시에, 실질적인 사교육 효과는 중간 또는 중간보다 약간 밑의 하위권에게만 나타난다는 것을 알 수 있다. 뿐만 아니라 선택편의의 통제 이후, 0.25분위와 0.5분위에 대한 계수 값도 절반 이하로 현저히 감소한 것으로 볼 때, 무조건부 분위회귀분석에서도 마찬가지로, 선택편의를 통제하지 않을 경우, 사교육 참여 효과에 대한 과대추정이 발생할 수 있다는 점을 염두에 두어야 한다.

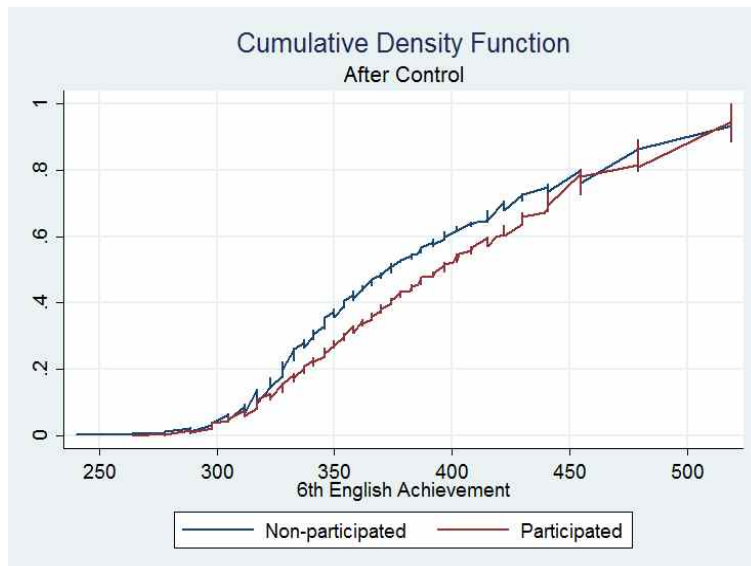
<표 IV-19> 영어교과 조건부 분위회귀분석 계수 비교

구분	0.1분위	0.25분위	0.5분위	0.75분위	0.9분위
6차 학업성취도					
통제 전	11***	31***	58***	47***	40***
통제 후	0	13***	23***	14	0
계수 차이	11	18	35	33	40

***p<.001, **p<.01, *p<.05



[그림 IV-5] 영어교과 선택편의 통제 전 누적확률분포



[그림 IV-6] 영어교과 선택편의 통제 후 누적확률분포

영어교과에서도 마찬가지로 선택편의 통제에 따라 사교육 참여 집단과 미참여집단의 누적 확률 분포(Cumulative Distribution Function)가 어떻게 달라지는지 [그림 IV-5]와 [그림 IV-6]에서 확인할 수 있다.

[그림 IV-5]와 [그림 IV-6]을 대조해보면 선택편의를 통제하기 전 두 누적확률 분포 사이의 거리는 선택편의를 통제한 후의 거리와 비교해 봤을 때 훨씬 그 폭이 넓다는 것을 알 수 있다. 또한 선택편의를 통제하기 전에는 거의 모든 분위에서 분포 사이의 폭이 넓었던 것에 비해, 선택편의를 통제하고 난 이후에 하위분위와 0.75이상의 상위 분위는 폭이 매우 좁거나 존재하지 않는다. 이는 곧 최하위권과 중상위권 분위에서 선택편의가 크게 발생하며 만약 통제하지 않은 채로 사교육 효과를 추정하는 경우 그 추정치가 과대추정 되었을 가능성을 암시한다.

5. 결론

1. 요약

이 연구의 목적은 분위회귀분석을 이용하여 사교육 효과를 추정할 때 선택편의의 통제에 따라 각 분위 효과의 유의도와 계수가 어떻게 달라지는지 밝히는 데 있다. 이를 위해 서울교육종단연구 6차년도 중학교 3학년 자료를 대상으로, 사교육 참여가 수학과 영어교과의 학업성취도에 미치는 분위별 영향을 추정하였다. 동시에 사교육 참여에 대한 선택편의를 통제한 후, 사교육 참여가 학업성취도에 미치는 영향을 다시 추정함으로써, 선택편의 통제 이전과 이후 사교육 참여의 분위 효과가 어떻게 달라지는지 비교하였다.

이에 따른 연구문제는 아래와 같다.

연구문제 1. 학업성취도에 대한 사교육 참여의 분위 효과를 분석할 때, 선택편의를 통제하지 않은 경우 각 분위 효과는 어떻게 나타나는가?

연구문제 1-1. 학업성취도에 대한 사교육 참여의 조건부 분위 효과를 분석할 때, 선택편의를 통제하지 않은 경우 각 분위 효과는 어떻게 나타나는가?

연구문제 1-2. 학업성취도에 대한 사교육 참여의 무조건부 분위 효과를 분석할 때, 선택편의를 통제하지 않은 경우 각 분위 효과는 어떻게 나타나는가?

연구문제 2. 학업성취도에 대한 사교육 참여의 분위 효과를 분석할 때, 선택편의를 통제한 경우 각 분위 효과는 어떻게 나타나

는가?

연구문제 2-1. 학업성취도에 대한 사교육 참여의 조건부 분위 효과를 분석할 때, 선택편의를 통제한 경우 각 분위 효과는 어떻게 나타나는가?

연구문제 2-2. 학업성취도에 대한 사교육 참여의 무조건부 분위 효과를 분석할 때, 선택편의를 통제한 경우 각 분위 효과는 어떻게 나타나는가?

이 연구에서는 사교육 참여에 대한 선택편의를 통제하기 위해, 경향점수 역확률 가중치(Inverse Probability Treatment Weight)를 사용하였다. 경향점수 역확률 가중치 방법은 개별 관측치의 배경 변수에 따라 사교육을 받을 확률을 산출하고, 그 확률의 역수를 해당 관측치에 대한 가중치로 사용하는 방법이다. 이 연구에서는 5차년도 자료로부터 얻은 사전 공변인을 사용하여 6차년도 사교육 참여에 대한 경향점수를 구하였으며 한 관측치에 과도하게 가중치가 높게 적용되는 것을 막기 위해 안정화 역확률 가중치를 사용하였다.

다음으로 선택편의 통제에 따른 사교육의 분위 효과 변화를 분석하기 위해, 먼저 조건부 분위회귀분석과 무조건부 분위회귀분석을 각각 실시하고, 각 분위에 대한 분위별 사교육 효과를 추정하였다. 이어서 안정화 역확률 가중치를 적용한 조건부 분위회귀분석과 무조건부 분위회귀분석을 각각 실시하여 선택편의가 통제된 분위별 사교육 효과를 추정하였다. 이 때, 조건부 분위회귀분석에는 학업성취도와 관련이 있는 것으로 알려진 변인 중 6차년도에 수집된 자료를 공변인으로 투입하였다.

연구의 결과를 요약하여 제시하면 다음과 같다.

먼저, 선택편의를 통제하기 전 조건부 분위회귀분석을 실시한 결과, 수학교과에 대한 사교육의 효과는 모든 조건부 분위에서 정적으로 유의하였고 분위가 높아질수록 효과의 크기가 증가하는 형태를 보여, 김민선과 백일우(2016)와 일관된 결과를 나타냈다. 한편, 사교육 참여에 대한 선택편의

를 통제하고 난 이후 사교육의 효과를 분석한 결과, 여전히 모든 조건부 분위에서 유의한 것으로 나타났다. 또한 그 크기가 분위가 높아질수록 증가하는 것으로 나타났다. 그러나 효과의 크기는 통제 이전과 비교했을 때 전 분위에서 상당 부분 감소하였으며, 감소의 폭은 분위가 높아질수록 증가하였다. 이는 사교육 참여에 대한 선택편의가 성적분위가 높아질수록 크게 발생한다는 것을 의미한다. 따라서 만약 선택편의를 통제하지 않고 사교육 참여의 효과를 분석할 경우 분위가 높아질수록 과대추정의 가능성이 높아질 수 있다.

둘째, 6차년도 공변인을 통제하지 않은 무조건부 분위회귀분석에서는, 선택편의 통제 이전에는 모든 분위에서 사교육 참여의 효과가 유의하였던 것에 비해, 통제 이후에는 0.9분위에서는 더 이상 유의하지 않음과 동시에 0.1분위에서는 t 값이 다소 감소하여, 최상위권과 최하위권에서는 사교육이 실질적 효과를 갖지 않는 것으로 나타났다. 또한 통제 후에도 여전히 효과가 유의한 것으로 나타난 0.25분위, 0.5분위, 0.75분위에서도 그 크기가 현저하게 감소하였으며, 이들 분위 내에서 감소의 폭은 분위가 높아질수록 증가하여, 무조건부의 경우, 주로 최상위권 분위에서 선택편의가 크게 발생함을 알 수 있다.

셋째, 영어교과에 대해 선택편의를 통제하지 않고 조건부 분위회귀분석을 실시한 결과, 모든 분위에서 사교육 참여가 정적으로 유의한 것으로 나타났다. 그러나 선택편의 통제 이후에는 상대적으로 하위분위인 0.1분위, 0.25분위와 중상위인 0.75분위에서 더 이상 유의하지 않았으며, 여전히 유의하게 나타난 0.5분위와 0.9분위에서는 t 값이 감소하였고, 그 크기 또한 감소하였다. 이를 통해 영어 교과에서도 사교육 참여에 대한 선택편의를 통제하지 않고, 사교육의 효과를 분석하는 경우, 과대추정의 가능성이 있으며 주로 하위분위에서 선택편의가 크게 작용하여, 실질적으로 사교육이 하위권에서 갖는 효과의 수준이 미미하다는 것을 알 수 있다.

넷째, 무조건부 분위회귀분석을 통해 사교육 효과를 추정한 결과, 선택편의 통제 이전에는 모든 분위에서 학업성취도에 정적으로 유의한 것으로

나타났으나 선택편의 통제 이후 최하위 분위인 0.1분위와 상위분위인 0.75 분위, 최상위분위인 0.9분위에서 더 이상 유의하지 않은 것으로 나타났다. 더불어 사교육 효과가 유의한 나머지 분위에서도 t값이 감소하고, 계수의 크기가 절반 이하로 작아지는 것을 확인할 수 있었다. 이를 통해 무조건부 분위에서는 상위분위에서 선택편의가 크게 작용하기 때문에 실질적으로 사교육이 상위권에 대해 갖는 효과의 수준이 미미하다는 것을 시사한다.¹⁶⁾

2. 논의

이 연구에서 나타난 결과를 종합해 보았을 때 다음과 같은 시사점을 얻을 수 있다.

첫째, 연구의 결과에 따르면 선택편의를 통제하기 전 나타나는 사교육의 분위별 효과는 선택편의를 통제한 후 나타나는 분위별 효과보다 대체로 더 유의하며 크기 또한 더 높게 추정되었다. 이는 분위회귀분석을 통해 사교육 효과를 추정하는 경우, 선택편의를 통제하지 않은 상태에서는 1종 오류나 과대추정의 가능성이 높아질 수 있다는 점을 시사한다. 즉, 사교육의 효과가 없거나 실제로는 효과의 크기가 낮음에도 불구하고, 사교육 참여에 대한 선택편의를 통제하지 않는 경우 효과가 있는 것으로 드러나거나 효과의 크기가 더 크게 산출될 수 있다. 이는 분위회귀분석에서도 단순회귀분석과 마찬가지로 처치효과를 산출하기 위해서는 처치집단과 미처치 집단 사이에 무선 할당을 가정하여야 한다는 이론과 상통한다(Liao and Zhao, 2007; Firpo, 2007; Fumagalli, 2009; Frolic & Melly, 2010). 일부에서는 분위회귀분석이 단순회귀분석에서는 추정할 수 없는 이질적 효과를 추정하고, 분위의 기준이 종속변수인 학업성취도에 있다는 점에서 사교육

16) 조건부 분위회귀분석에서 선택편의가 크게 발생하는 분위와, 무조건부 분위회귀분석에서 선택편의가 크게 발생하는 분위 다른 것은 전자는 공변인이 통제된 상태에서의 성적 분위기를, 후자는 공변인이 통제되지 않은 상태에서의 분위기를 말하기 때문이다. 이에 대한 자세한 설명은 15쪽에서 확인할 수 있다.

참여에 대한 선택편의를 고려하지 않을 수 있겠으나, 이질적 효과를 분석하는 것과 선택편의를 통제하는 것은 별개의 문제로 보아야 한다.

이 때, 사교육 효과에 대한 분위회귀분석에서 선택편의가 발생하는 이유는 다음과 같다. 먼저, 분위회귀분석에서의 효과란 처치집단과 미처치집단 간의 분위수 차이를 의미하는데, 사실 사교육 참여 집단과 미참여 집단 간에 특정 분위에서의 점수 차이는 단순히 사교육 참여의 효과가 아니라, 애초에 사교육 참여집단과 미참여 집단의 특성이 다르기 때문에 비롯되었을 수 있다. 특히, 이 연구의 결과로부터 최상위 또는 상위 분위의 선택편의가 두드러진다는 것을 알 수 있었는데 이는 전체 학생들의 6차년도 성적 분포를 단일한 분포로 가정할 때, 이 분포에서 중상위권의 학생들은 사교육을 받는 학생과 받지 않는 학생의 비율이 어느 정도 균형있게 나뉘어 있는 반면, 최상위권 학생들의 경우 거의 모든 학생들이 사교육을 받아 비롯된 것으로 생각할 수 있다. 이 경우, 대부분의 최상위권 또는 상위권 학생들이 참여집단으로 편성된 것으로, 참여집단의 최상위권 점수와 미참여 집단 간의 최상위권 점수의 차이가 두드러진 것으로 보인다.

둘째, 선택편의 통제 전후 사교육 효과의 추정치는 대체로 줄어들거나 유의해지지 않는 방향으로 바뀌었지만 분포의 형태, 분위, 교과에 따라 조금씩 다른 양상을 보였다. 이는 통제 조건 별로, 성적 분위별로, 각 교과별로 선택편의가 다르게 나타난다는 것을 암시한다. 예를 들어, 수학교과와 영어교과를 비교할 때 해당연도 배경변인이 통제된 조건부 분포에서는 선택편의 통제가 계수의 유의성에는 영향을 주지 못하였으나, 효과 크기는 분위가 높을수록 감소의 폭이 증가하였다. 그러나 무조건 분포에 대해서는 선택편의를 통제하였을 경우 분위가 높을수록 계수의 크기는 크게 감소하였으나 최상위권에서 사교육의 효과는 더 이상 유의하지 않은 것으로 나타났다. 마찬가지로 영어교과에서는 수학교과와 다소 다른 양상을 보였는데, 수학 교과의 경우 선택편의의 통제가 주로 상위분위와 하위분위의 효과에 많은 영향을 미치는 것으로 나타난 반면, 영어 교과의 경우 상대적으로 중간에 가까운 분위까지도 유의도 변화에 영향을 미쳤다. 이는 영어 교과의 경우 수학 교과에 비해 더 많은 부분에서 선택편의가 발생하고 있음을 암시한다. 상기

의 결과로 미루어 볼 때 사교육 효과 연구에 있어 선택편의는 모든 분포, 분위, 교과에서 동일하게 발생하는 것이 아니라 각각의 세부적인 조건에 따라 다르게 발생한다는 것을 알 수 있다.

이러한 연구 결과를 바탕으로 다음과 같은 후속연구를 제안할 수 있다.

먼저, 이 연구에서는 조건부 분위회귀분석과 무조건부 분위회귀분석에 대하여 역확률 가중치를 적용하여 선택편의를 통제하였는데, 사실 이 방법은 처치의 선택이 관찰 가능한 변인에 의해 충분히 예측될 수 있다는 가정이 전제되어 있다. 그러나 사교육의 선택이 관찰 가능하지 않은 변인에 의해서도 예측될 수 있다는 가능성을 배제할 수 없기 때문에 Abadie, Angrist와 Imbens(2002)나 Frolich와 Melly(2008)와 같이 도구변수를 활용한 방법으로 선택편의의 영향을 확인해 볼 필요가 있다.

둘째로 이 연구는 서울교육중단연구 중학생 자료를 대상으로 분석을 실시하였다. 이 연구에서 실증데이터 분석을 선택한 것은 분위회귀분석에 있어 처치효과 추정을 위해서는 처치에 대한 무작위성이 전제되어야 하고, 이에 따라 선택편의를 통제하는 것이 실제 효과 추정치에 더 근접할 것이라는 이론적 가정을 하고 있기 때문이다. 그러나 데이터의 특성 별로 선택편의의 발생이 다르게 나타날 수 있고, 때에 따라 선택편의 통제의 영향이 미미할 가능성이 있다. 따라서 시뮬레이션 연구를 통해 어떤 특성의 표본 집단에서 선택편의 통제의 영향이 어느 정도 나타나는지를 밝히면, 각 표본 집단별로 선택편의를 다르게 잡아 주어 보다 정확한 사교육 효과 추정이 가능해 질 것으로 보인다.

<참 고 문 헌>

- 강창희(2012). ‘학교교육 수준 및 실태 분석 연구: 중학교’ 자료를 이용한 사교육비 지출의 성적 향상효과 분석. 한국개발연구, 34(2), 139-171.
- 강창희, 이정민, 이석배, 김세움(2013). 관광정책 및 관광사업 프로그램 평가방법. 문화체육관광부.
- 길혜지, 백순근(2016). 성적분위와 거주 지역별 EBS 수능강의 수강이 수능점수에 미치는 차별적 효과 분석. 교육평가연구, 29(3), 543-563.
- 김민선, 백일우(2016). 사교육에 대한 시간 및 비용 투자에 따른 효과 분석: 서울 중학생의 학업성취도 분위별 효과를 중심으로. 교육재정경제연구, 25(3), 1-26.
- 김병철(2010). 사교육의 경제적 효과 -대학진학과 노동시장 성과를 중심으로-. 서울대학교 대학원 석사학위논문.
- 김성식, 송혜정(2013). 학교특성에 따른 사교육 참여와 효과 차이 분석. 교육평가연구, 26(5), 1187-1215.
- 김현진(2007). 가정배경, 학교교육, 그리고 사교육이 학업성취에 미치는 영향 분석. 교육행정학연구, 25(4), 485-508.
- 김희삼(2010). 학업성취도, 진학 및 노동시장 성과에 대한 사교육의 효과 분석. 서울: 한국개발연구원.
- 남기곤(2008). 사교육시간과 학업성적과의 관련성: PISA 자료를 이용한 국제비교 분석. 한국경제학보, 15(1), 55-90.
- 박창남, 도종수(2005). 부모의 사회경제적 지위가 학업성취도에 미치는 영향. 사회복지정책, 22, 281-303.
- 박현정(2010). 학생들의 사교육 참여와 수학성취도 및 수학 수업이해도 간 관계에 대한 종단적 분석. 교육평가연구, 23(4), 887-907.
- 박현정, 하여진, 박민호, 오현철, 이준호, 이진실(2011). 서울교육종단연구 2010 사용자 매뉴얼. 서울특별시교육연구정보원.

- 상경아, 백순근(2005). 고등학생의 수학 과외가 학업성취도, 태도, 자기조절학습에 미치는 영향. 교육평가연구, 18(3), 39-57.
- 상경아(2009). 경향점수를 이용한 결합표집 방법에 의한 사교육 효과 분석. 교육평가연구, 22(3), 717-735.
- 성기선, 김준엽(2010). 고등학생들의 사교육 참여 시간과 학업성취도와의 관련성 비교연구 -한국, 일본, 핀란드 과학성취도를 중심으로-, 20(1), 103-126.
- 양정호(2003). 중학생의 과외참여 요인에 관한 연구: TIMSS-R의 위계적 일반화선형모형 분석. 한국교육, 30(2), 261-283.
- 양정호(2006). 한국의 사교육비 격차 추세에 관한 연구: 한국노동패널조사의 다극화 지수와 지니 계수를 이용한 분석. 교육재정경제연구, 15(2), 199-219.
- 오영수, 윤정식(2003). 일반계 고등학생의 성적결정 요인 분석. 교육재정경제연구, 12(1), 33-56.
- 윤유진, 김현철(2016). 사교육, 방과후학교, EBS 참여가 학업성취에 미치는 효과분석. 교육행정학연구, 26(34), 385-417.
- 윤정일 (1997). 초, 중등 과외실태 조사연구. 교육재정경제연구, 6(3), 231-254.
- 이광현, 권용재(2011). 사교육비와 사교육시간이 학업성취도에 미치는 효과 분석: 분위회귀분석을 이용한 접근. 교육재정경제연구, 20(3), 99-133.
- 이석원(2006). 실업자 직업훈련 사업의 분포적 효과: 분위수 회귀식을 사용한 분석. 행정논총, 44(4), 149-176.
- 이수정, 임현정(2009). 중학생의 학업성취에 대한 사교육비 효과 분석. 교육재정경제연구, 18(1), 141-166.
- 임연기(1997). 초, 중등학생의 과외수업 실태와 정책적 시사점. 교육재정경제연구, 6(3), 317-360.
- 임천순, 박소영, 이광호(2004). 사교육이 학업성취에 미치는 영향, 교육재정경제연구, 13(1), 331-356.
- 정동욱, 박현정, 하여진, 박민호(2012). EBS 교육 프로그램의 사교육 경감 효과 분석. 교육행정학연구, 30(3), 21-42.

- 차동춘(2015). 중학교 수학 학업 성취도 성장에 대한 사교육의 효과. *교원교육*, 31(3), 247-272.
- 통계청(2017). 2016년 초·중·고 사교육비조사 결과 보도자료. 2017년 3월 14일자.
- 하여진, 박현정(2015). 다중처치의 선택편의를 조정 한 EBS 수능강의 효과 분석. *교육평가연구*, 28(3), 831-852.
- 한대동, 성병창, 길임주(2001). 고등학생 학업성취에 대한 학교효과와 과외효과의 비교연구. *교육사회학연구*, 11(1), 33-54.
- Abadie, A., J. Angrist, and G. Imbens. (1998). Instrumental Variables Estimation of Quantile Treatment Effects. Technical Working Paper Series, Technical Working Paper 229, National Bureau of Economic Research.
- Abadie, A., J. Angrist, and G. Imbens. (2002). Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings. *Econometrica*, 70, 91-117.
- Angrist, J., and J. Pischke. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. NJ: Princeton University Press.
- Buchinsky, M. (1998). Recent Advances in Quantile Regression Model: A Practical Guideline for Empirical Research, *The Journal of Human Resources*, 33(1), 88-126.
- Buchinsky, M. (2001). Quantile Regression with Sample Selection: Estimating Women's Return to Education in the U.S.. *Empirical Econometrics*, 26, 87-113.
- Bitler, P., and J. Gelbach. and W. Hoynes. (2003). What Mean Impact Miss: Distributional Effects of Welfare Reform Experiments. NBER Working Paper Series, Working Paper 10121, National Bureau of Economic Research.
- Caliendo, M., and S. Kopeinig. (2005). Some Practical Guidance for the Implementation of Propensity Score Matching. DIW Discussion Papers, No. 485.

- Chernozhukov, V. and C. Hansen. (2005). An IV Model of Quantile Treatment Effects. *Econometrica*, 73, 245-261.
- Doksum, K. (1974). Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-sample Case. *The Annals of Statistics*, 2(2), 267-277.
- Firpo, S. (2007). Efficient Semiparametric Estimation of Quantile Treatment Effects. *Econometrica* 75(1), 259-276.
- Firpo, S., Fortin, N. M. and Lemieux, T. (2007). Unconditional Quantile Regressions. NBER Technical Working Papers 0339, National Bureau of Economic Research, Inc.
- Firpo, S., Fortin, N. M. and Lemieux, T. (2009). Unconditional Quantile Regressions. *Econometrica*, 77(3), 953-973.
- Frolich, M. and B. Melly. (2010). Unconditional Quantile Treatment Effects under Endogeneity. Discussion Paper No.3288, Institute for the Study of Labor(IZA). <http://ideas.repec.org/p/iza/izadps/dp3288.html>.
- Frolich, M. and B. Melly. (2010). Estimation of Quantile Treatment Effects with Stata. *The Stata Journal*, 10(3), 423-457.
- Fumagalli, E. (2012). Average and Quantile Treatment Effects of the American Folic Acid Fortification: an evaluation in a quasi-experimental framework. University of York Health, Econometrics and Data Group (HEDG) Working Paper 12/08
- Heckman, J., J. Smith, and N. Clements. (1993). Making the Most Out of Programme Evaluations and Social Experiments: Accounting For Heterogeneity in Programme Impacts. *Review of Economic Studies*, 64, 487-535.
- Kang, C. (2007). "Does Money Matter? The Effect of Private Educational Expenditures on Academic Performance." Departmental Working Papers from National University of Singapore, Department of Economics.

- Killewald A. and J. Bearak. (2010). Is the Motherhood Penalty Larger for Low-Wage Women? A Comment on Quantile Regression. *American Sociological Review*, 79(2), 350-357.
- Koenker, R., and G. Bassett. (1978). Regression Quantiles, *Econometrica*, 46, 33-50.
- Koenker, R., and K. Hallock. (2000). Quantile Regression: An Introduction. Unpublished Paper.
- Koenker, R. (2005). Quantile Regression. NY: Cambridge University Press.
- Lehmann, E. (1974). Nonparametrics: Statistical Methods Based on Ranks. CA: Holden-Day.
- Liao, W.-C., Zhao, D. (2017). The Selection and Quantile Treatment Effects on the Economic Returns of Green Buildings. working paper.
- Rajeev, H. D., & Sadek, W. (2002). Propensity Score-Matching Methods for Nonexperimental Causal Studies. *The Review of Economics and Statistics*, 84(1), 151-161.
- Rangvid B. S. (2003). Educational Peer Effects. Quantile Regression Evidence from Denmark with PISA2000 data. Doctoral Dissertation. Aarhus School of Business.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70, 41-55.
- Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 56, 688-701.
- Xu, S., Ross, C., Raebel, M. A., Shetterly, S., Blanchett, C., & Smith, D. (2010). Use of Stabilized Inverse Propensity Scores as Weight to Directly Estimate Relative Risk and Its Confidence Intervals. *Value In Health*, 13(2), 273-277.

Abstract

Analysis of Quantile Effects of Private Education on Academic Achievement and Control of Sample Selection

Kim Ji Hye

Department of Education

The Graduate School

Seoul National University

The purpose of this research is to figure out how the significance and coefficients in each quantile effects vary depending on the control of the selection bias, when estimating the effects of the private tutoring using the quantile regression analysis. For this purpose, the effects of participation of private tutoring on the academic performance of Mathematics and English in the 3rd year of middle school have been analyzed, using 6th Seoul Educational Longitudinal Study of 2010. At the same time, this research paper also presents how the control of selection bias affects the entire result by comparing the

results of pre and post control of selection bias for participation of private tutoring.

As such, the research questions are listed as the below.

Question 1. When analyzing the quantile effects of participation in private tutoring, how does each quantile effect appeared when the selection bias is not controlled?

Question 2. When analyzing the quantile effects of participation in private tutoring, how does each quantile effect get appeared when the selection bias is controlled?

This research uses Inverse Probability Treatment Weight (IPTW) to control the selection bias for participation in private tutoring. The method calculates the probability of participating in private tutoring, based on the background variables of individual observations, and uses the reciprocal of the probabilities as the weight for each observations. The above method is applied in this research by using the probabilities for participation in private tutoring in the 6th year, which is anticipated from the covariates in the 5th year data. In addition, to prevent excessive weighting of a single observation stabilized inverse probabilities treatment weight is also applied. Upon selection of the method and the subject, to analyze the changes in the quantile effects of private education depending on control of selection bias, this research separately performs Conditional Quantile Regression(CQR) analysis and Unconditional Quantile Regression(UQR) analysis to estimate quantile effect for each distributions. Next, stabilized inverse probabilities treatment weights are applied to each conditional quantile regression and unconditional quantile regression for the estimation of the quantile effects of private education, under control of selection bias. Specifically, for the conditional quantile regression analysis

variables from data collected in the 6th year, which are known to be related to academic achievement, have been used as covariates.

The results and implications of this research are summarized as follows.

Firstly, before controlling the selection bias, the quantile effects of private tutoring appeared to be more significant and larger than the quantile effects after controlling the selection bias. This implies that in the case of estimating the effect of private tutoring through quantile regression analysis without controlling the selection bias, the probability of type 1 error or overestimation gets higher. In other words, even when the effect of private tutoring is not significant or the magnitude of the effect is small in reality, the result may appear to be effective. This is consistent with the theory that, like simple regression analysis, random assignment between the treatment group and the un-treatment group shall be assumed for the estimation of the treatment effect (Abadie, Angrist, & Imbens, 2002; Chernozhukov & Hansen, 2005; Firpo, 2007; Frolich & Melly, 2008).

Secondly, the estimates of the effects of private tutoring before and after the control of selection bias, were changed in a direction, generally less or not significant, but showed a slightly different patterns depending on the distribution type, quantiles and subjects. This implies that the selection bias take place differently for each condition, based on control, achievement quantiles and subjects. In particular, it is found that selection bias of the highest quantile is prominent, because the difference in the highest scores between the two groups is remarkable, as the majority of the top level students were organized into participating group. For mathematics, the control of selection bias does

not affect significance of the coefficient in the conditional distribution, but the magnitude of effect decrease notably with the increase of the quantile, and also, in the unconditional distribution, the coefficient of magnitude decreased greatly and even become no more significant in the highest quantile when selection bias controlled. Moreover, for the unconditional distribution, the coefficient of magnitude decreases greatly and even becomes no more significant in the highest quantile when selection bias controlled. On the other side, English shows a quite different results, presenting that selection bias impact even the middle quantiles including upper and lower quantiles, while for mathematics, it is presented that control of selection bias affected mainly the upper and lower quantiles only. This suggests that selection bias in English affects in larger areas than it does in Mathematics. In conclusion, these results suggest that the selection bias in the study of private tutoring effects does not occur evenly in all distributions, quantiles and subjects, but varies depending to each detailed conditions.

keywords : Private Education, Quantile Regression, Selection Bias, Propensity Score, Inverse Probability Treatment Weight, Quantile Treatment Effect

Student Number : 2016-21503

부 록

<안정화 역확률 가중치 산출 코드>

*로짓분석 실시 결과

```
logit FP29M ES09M ES10M ES13CM ES22 ES54 EP08_1 EP16 EP29M  
ES00BM AP05  
predict preprob, pr
```

*역확률 가중치 산출 및 확인

```
gen pweight=.  
replace pweight=(1/preprob) if FP29M==1  
replace pweight=(1/(1-preprob)) if FP29M==0  
summarize pweight, detail
```

*안정화 역확률 가중치 산출 및 확인

```
logit FP29M  
predict baseprob, pr  
  
gen sweight_tr=.  
replace sweight_tr=baseprob/preprob if FP29M==1  
replace sweight_tr=(1-baseprob)/(1-preprob) if FP29M==0
```

<단순회귀분석 및 조건부 분위회귀분석 코드>

*OLS and OLS with IPTW

```
reg FS00BM FP29M FS09M FS10M FS13CM FS27 FS39 AP05 DC99A DS01
reg FS00BM FP29M FS09M FS10M FS13CM FS27 FS39 AP05 DC99A DS01
[pweight=sweight]
```

*Conditional Quantile Regression without IPTW

```
qreg FS00BM FP29M FS09M FS10M FS13CM FS27 FS39 AP05 DC99A DS01,
quantile(.10) vce(robust)
qreg FS00BM FP29M FS09M FS10M FS13CM FS27 FS39 AP05 DC99A DS01,
quantile(.25) vce(robust)
qreg FS00BM FP29M FS09M FS10M FS13CM FS27 FS39 AP05 DC99A DS01,
quantile(.50) vce(robust)
qreg FS00BM FP29M FS09M FS10M FS13CM FS27 FS39 AP05 DC99A DS01,
quantile(.75) vce(robust)
qreg FS00BM FP29M FS09M FS10M FS13CM FS27 FS39 AP05 DC99A DS01,
quantile(.90) vce(robust)
```

*Conditional Quantile Regression with IPTW

```
qreg FS00BM FP29M FS09M FS10M FS13CM FS27 FS39 AP05 DC99A DS01,
quantile(.10) vce(robust), [pweight=sweight_tr]
qreg FS00BM FP29M FS09M FS10M FS13CM FS27 FS39 AP05 DC99A DS01,
quantile(.25) vce(robust), [pweight=sweight_tr]
qreg FS00BM FP29M FS09M FS10M FS13CM FS27 FS39 AP05 DC99A DS01,
quantile(.50) vce(robust), [pweight=sweight_tr]
qreg FS00BM FP29M FS09M FS10M FS13CM FS27 FS39 AP05 DC99A DS01,
quantile(.75) vce(robust), [pweight=sweight_tr]
qreg FS00BM FP29M FS09M FS10M FS13CM FS27 FS39 AP05 DC99A DS01,
quantile(.90) vce(robust), [pweight=sweight_tr]
```

<무조건부 분위회귀분석 코드>

*Unconditional Quantile Estimation

```
qreg FS00BM FP29M, quantile(.10), [pweight=sweight_tr]
qreg FS00BM FP29M, quantile(.25), [pweight=sweight_tr]
qreg FS00BM FP29M, quantile(.50), [pweight=sweight_tr]
qreg FS00BM FP29M, quantile(.75), [pweight=sweight_tr]
qreg FS00BM FP29M, quantile(.90), [pweight=sweight_tr]
```

*Unconditional Quantile Estimation with IPTW

```
qreg FS00BM FP29M, quantile(.10), [pweight=sweight_tr]
qreg FS00BM FP29M, quantile(.25), [pweight=sweight_tr]
qreg FS00BM FP29M, quantile(.50), [pweight=sweight_tr]
qreg FS00BM FP29M, quantile(.75), [pweight=sweight_tr]
qreg FS00BM FP29M, quantile(.90), [pweight=sweight_tr]
```

*ivqte package 이용하여 Unconditional Quantile Treatment Estimation

```
ivqte FS00BM (FP29M), variance quantiles(0.1, 0.25, 0.50, 0.75, 0.90)
continuous(ES09M ES10M ES13CM ES22 ES54 EP08_1 EP16 EP29M ES00BM
AP05)
```